



# INTELIGENCIA DE NEGOCIO

2020 - 2021

- Tema 1. Introducción a la Inteligencia de Negocio
- Tema 2. Minería de Datos. Ciencia de Datos
- Tema 3. Modelos de Predicción: Clasificación, regresión y series temporales
- Tema 4. Preparación de Datos
- Tema 5. Modelos de Agrupamiento o Segmentación
- Tema 6. Modelos de Asociación
- Tema 7. Modelos Avanzados de Minería de Datos.
- Tema 8. Big Data

# Inteligencia de Negocio

## TEMA 3. Modelos de Predicción: Clasificación, regresión y series temporales

---

1. Clasificación
- 2. Regresión**
3. Series Temporales



## Objetivos. Parte II. Regresión

---

- Entender qué es un problema de regresión
- Conocer los criterios que se utilizan para evaluar un algoritmo de regresión
- Conocer algoritmos concretos de regresión

# Inteligencia de Negocio

## TEMA 3. Modelos de Predicción: Clasificación, regresión y series temporales

---

### Regresión

(Numeric prediction, Section 6.5, pp 243, Witten and Frank, 2005, Data mining, Elsevier)

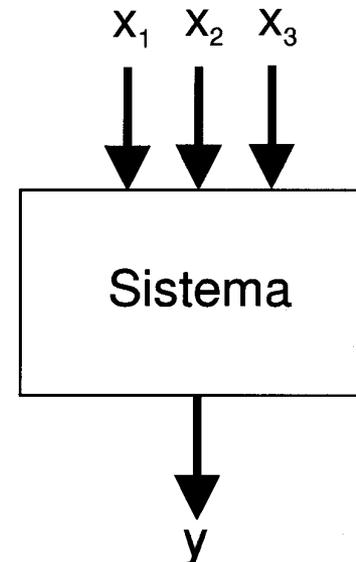
1. Definición del problema
2. Evaluación de algoritmos de regresión
3. Técnicas de clasificación válidas para regresión
4. Análisis de regresión
5. Árboles de regresión
6. Árboles de modelos
7. Sistemas basados en reglas difusas

# 1. Definición del problema

---

- Objetivo: predecir el valor numérico para una variable a partir de los valores para otras
- La definición del problema es parecida a la del problema de clasificación: tenemos variables predictoras y una variable de regresión que en este caso es numérica
- En regresión la mayoría (e incluso todas) las variables predictoras son numéricas

El problema fundamental de la predicción está en modelar la relación entre las variables de estado para obtener el valor de la variable a predecir.



# 1. Definición del problema

---

- Objetivo: predecir el valor numérico para una variable a partir de los valores para otras
- La definición del problema es parecida a la del problema de clasificación: tenemos variables predictoras y una variable de regresión que en este caso es numérica
- En regresión la mayoría (e incluso todas) las variables predictoras son numéricas
- **Ejemplos:**
  - **¿qué consumo tendrá un coche en autovía en función de su peso, cilindrada, potencia,...?**
  - **¿qué número de artículos tendremos para el próximo pedido?**
  - **¿cuántos meses necesitaremos para desarrollar un proyecto software?**
  - **¿cuál es la probabilidad de que un cliente determinado sea receptivo a un envío publicitario?**
  - **¿cuántos enfermos tendremos en urgencias la próxima nochebuena?**

## 2. Evaluación de algoritmos de regresión

---

- Todas las técnicas de validación estudiadas en clasificación son válidas para predicción numérica
- La diferencia está en que ahora debemos medir el error de otra forma
- Debemos medir el error cometido al aproximar un conjunto de valores  $\{v_1, \dots, v_n\}$  por su estimación  $\{v'_1, \dots, v'_n\}$

Error cuadrático medio (ECM)

$$ECM = \frac{\sum_{i=1}^n (v_i - v'_i)^2}{n}$$

ECM estandarizado (ECME)

$$ECME = \sqrt{\frac{\sum_{i=1}^n (v_i - v'_i)^2}{n}}$$

Error medio absoluto (EMA)

$$EMA = \frac{\sum_{i=1}^n |v_i - v'_i|}{n}$$

Error absoluto relativo (EAR)

$$EAR = \frac{\sum_{i=1}^n |v_i - v'_i|}{\sum_{i=1}^n |v_i - \bar{v}|}$$

Coeficiente de correlación  $r_{vv'} = \frac{\sum_{i=1}^n (v_i - \bar{v})(v'_i - \bar{v}')}{(n-1)\sigma_v\sigma_{v'}}$

# Inteligencia de Negocio

## TEMA 3. Modelos de Predicción: Clasificación, regresión y series temporales

---

### Regresión (Predicción numérica)

1. Definición del problema
2. Validación de algoritmos de regresión
3. Técnicas de clasificación válidas para regresión
4. Análisis de regresión
5. Árboles de regresión
6. Árboles de modelos
7. Sistemas basados en reglas difusas

# 3. Técnicas de clasificación válidas para regresión: kNN, RNN

---

- **Métodos basados en ejemplos/instancias:**

Al utilizar kNN, si los  $k$  vecinos más próximos  $\{e_1, \dots, e_k\}$  tienen valores  $\{v_1, \dots, v_k\}$  para la variable objetivo, entonces el valor a devolver para el objeto analizado  $e'$  sería

$$v = \begin{cases} \frac{\sum_{i=1}^k v_i}{k} & \text{si todos cuentan igual} \\ \frac{\sum_{i=1}^k w_i \cdot v_i}{\sum_{i=1}^k w_i} & \text{si se hace un voto ponderado} \end{cases}$$

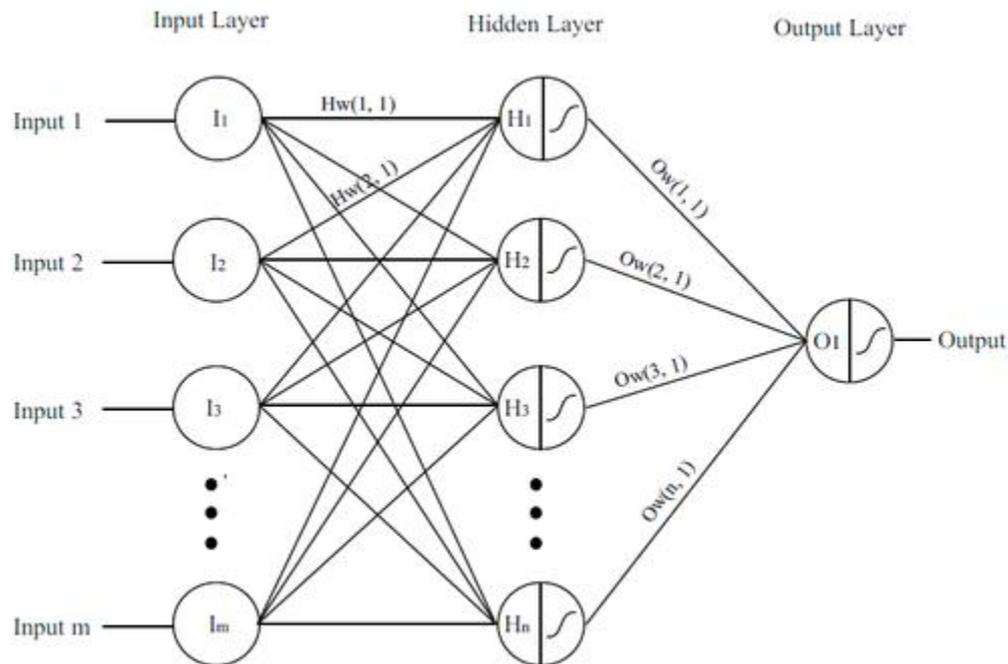
por ejemplo, con  $w_i = 1/d(e_i, e')$

# 3. Técnicas de clasificación válidas para regresión: kNN, RNN

---

## ■ Métodos basados en redes neuronales:

- La capa de salida sería una única neurona
- Como los pesos se adaptan en función del error cometido, es suficiente con medir de forma adecuada el error



# Inteligencia de Negocio

## TEMA 3. Modelos de Predicción: Clasificación, regresión y series temporales

---

### Regresión (Predicción numérica)

1. Definición del problema
2. Evaluación de algoritmos de regresión
3. Técnicas de clasificación válidas para regresión
4. Análisis de regresión
5. Árboles de regresión
6. Árboles de modelos
7. Sistemas basados en reglas difusas

# 4. Análisis de regresión

---

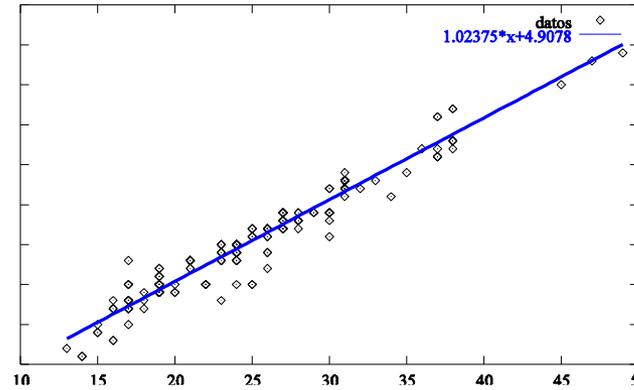
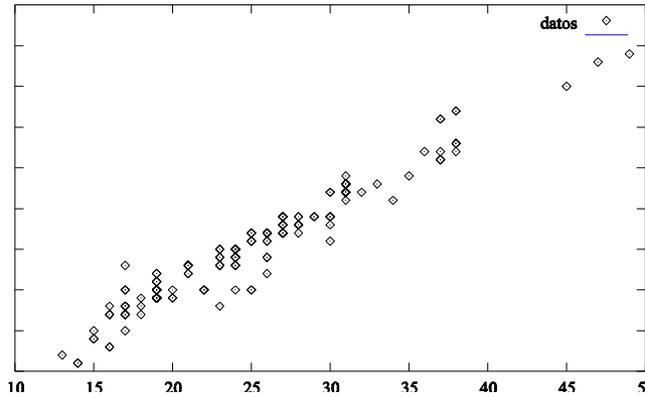
- El análisis de regresión es el método más utilizado para realizar la tarea de predicción numérica
- **Objetivo:** estimar la variable objetivo ( $y$ ) como una ecuación que contiene como incógnitas al resto de las variables ( $x_1, \dots, x_n$ )
- El modelo más sencillo es la **regresión lineal** que reducida a una sola variable predictora tiene la forma:
$$y = a + b \cdot x$$
- Estos coeficientes pueden obtenerse fácilmente mediante el método de los mínimos cuadrados

$$b = \frac{\sum_{i=1}^s (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^s (x_i - \bar{x})^2}$$

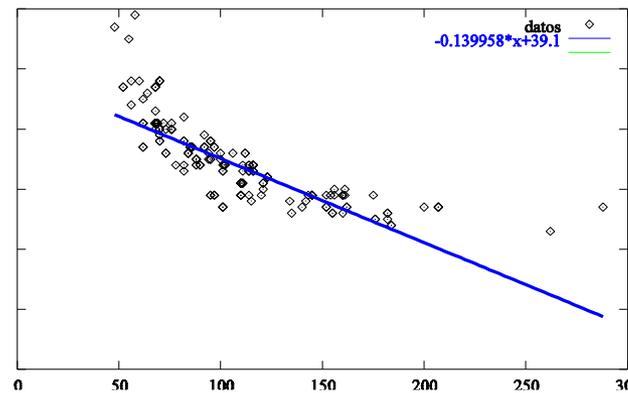
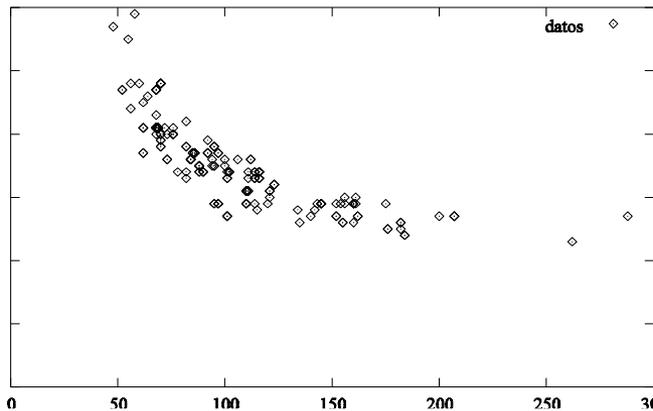
$$a = \bar{y} - b \cdot \bar{x}$$

# 4. Análisis de regresión

---



Bastante apropiado



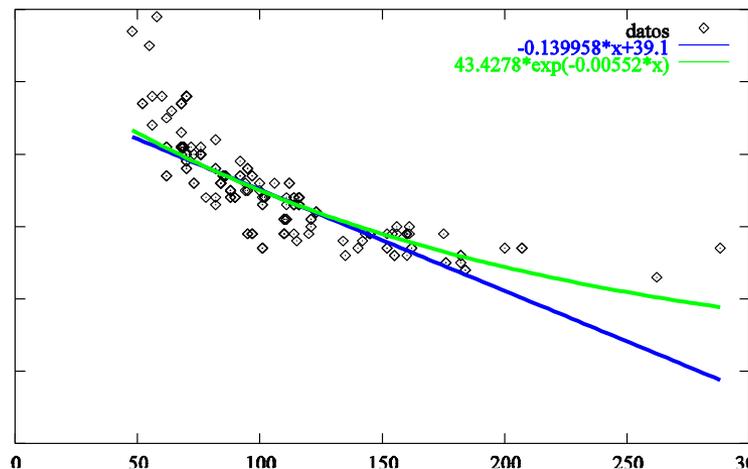
No tan apropiado

# 4. Análisis de regresión

- Para estimar curvas es necesario utilizar otra regresión, por ejemplo, **regresión exponencial**:  $y = a \cdot e^{bx}$
- ¿Cómo estimamos ahora a y b? Tomando logaritmos

$$\ln(y) = \ln(a \cdot e^{bx}) \Rightarrow \ln(y) = \ln(a) + \ln(e^{bx}) \Rightarrow y^* = a^* + bx$$

- Es decir, tenemos un problema de regresión lineal entre  $y^* = \ln(y)$  y  $x$ . Una vez estimados  $a^*$  y  $b$  podemos calcular  $a = e^{a^*}$



# 4. Análisis de regresión

---

## Regresión lineal múltiple

- Cuando hay más de una variable predictora, la ecuación de predicción se transforma en

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

Este problema se conoce como regresión lineal múltiple

- La estimación de los coeficientes es algo más compleja y requiere operar con matrices (y sus inversas)
- Por ejemplo, continuando con el ejemplo CPU.arff, tendríamos

$$\text{Class} = 0.066 * \text{MYCT} + 0.0143 * \text{MMIN} + 0.0066 * \text{MMAX} + 0.4945 * \text{CACH} - 0.1723 * \text{CHMIN} + 1.2012 * \text{HMAX} - 66.4814$$

$$\text{EAM} = 34.31 \%$$

$$\text{EAR} = 39.26 \%$$

- Existen técnicas más complejas de regresión que aproximan de forma más precisa los datos de entrada

**Regresión no-lineal, regresión logística, ...**

# Inteligencia de Negocio

## TEMA 3. Modelos de Predicción: Clasificación, regresión y series temporales

---

### Regresión (Predicción numérica)

1. Definición del problema
2. Evaluación de algoritmos de regresión
3. Técnicas de clasificación válidas para regresión
4. Análisis de regresión
5. Árboles de regresión
6. Árboles de modelos
7. Sistemas basados en reglas difusas

# 5. Árboles de regresión

---

- Un **árbol de regresión** es un árbol de decisión cuyas hojas predicen una cantidad numérica
- Ese valor numérico se calcula como la media del valor para la variable clase de todos los ejemplos que han llegado a esa hoja durante el proceso de construcción del árbol
- La evaluación de un nuevo ejemplo es idéntico a los árboles de decisión
- Durante el proceso de predicción es posible utilizar un suavizado de los valores del ejemplo a tratar, con el fin de salvar las posibles discontinuidades presentes en los datos
- El criterio de selección de una variable en la construcción del árbol está basado en una reducción del error esperado: reducción de la desviación/varianza en la variable objetivo (standard reduction deviation):

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} \cdot sd(T_i)$$

*SDR=Standard Deviation Reduction*

- Al final el árbol se poda para evitar el sobreajuste

# 5. Árboles de regresión

---

```
MMAX <= 14000 :
| CACH <= 8.5 :
| | MMAX <= 6100 : LM1 (75/5.4%)
| | MMAX > 6100 :
| | | MYCT <= 83.5 :
| | | | MMAX <= 10000 : LM2 (8/4.5%)
| | | | MMAX > 10000 : LM3 (3/3.78%)
| | | MYCT > 83.5 : LM4 (22/6.73%)
| CACH > 8.5 :
| | CHMIN <= 7 :
| | | MYCT <= 95 : LM5 (7/13.3%)
| | | MYCT > 95 :
| | | | CACH <= 28 : LM6 (12/7.85%)
| | | | CACH > 28 : LM7 (6/7.27%)
| | CHMIN > 7 : LM8 (8/45%)
MMAX > 14000 :
| MMAX <= 22500 :
| | CACH <= 27 :
| | | CHMIN <= 5 : LM9 (14/6.92%)
| | | CHMIN > 5 : LM10 (5/12.8%)
| | CACH > 27 : LM11 (18/25.1%)
LM1: class = 24.3
LM2: class = 45.3
LM3: class = 62.7
LM4: class = 39.6
LM5: class = 63.6
LM6: class = 43.9
LM7: class = 55
LM8: class = 104
LM9: class = 75.4
LM10: class = 94.4
LM11: class = 128
LM12: class = 262
LM13: class = 174
LM14: class = 355
LM15: class = 971
EAM 18.1275
EAR 20.7449%
WEKA: sin suavizar
poda=1
```

# Inteligencia de Negocio

## TEMA 3. Modelos de Predicción: Clasificación, regresión y series temporales

---

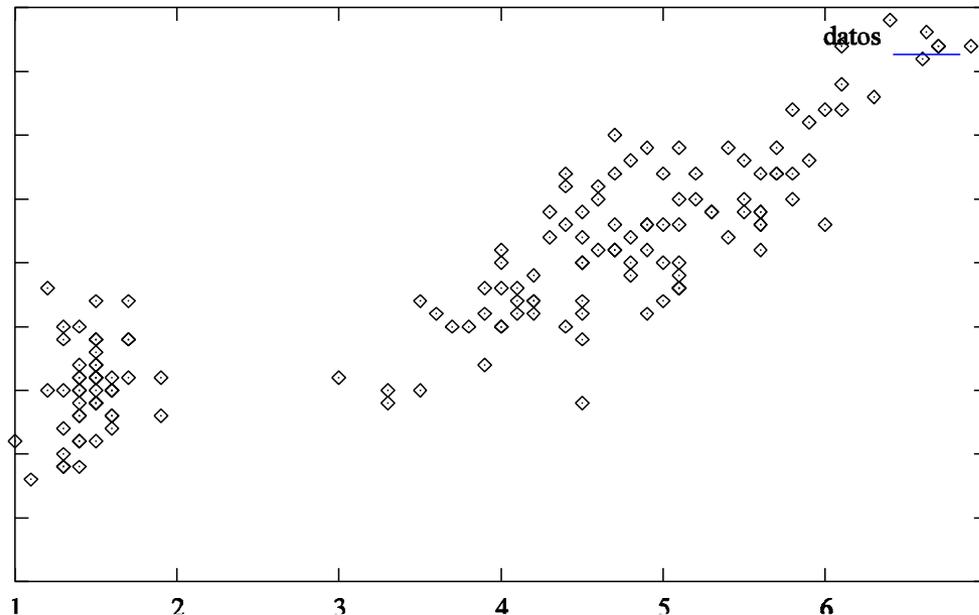
### Regresión (Predicción numérica)

1. Definición del problema
2. Evaluación de algoritmos de regresión
3. Técnicas de clasificación válidas para regresión
4. Análisis de regresión
5. Árboles de regresión
6. Árboles de modelos
7. Sistemas basados en reglas difusas

# 6. Árboles de modelos

---

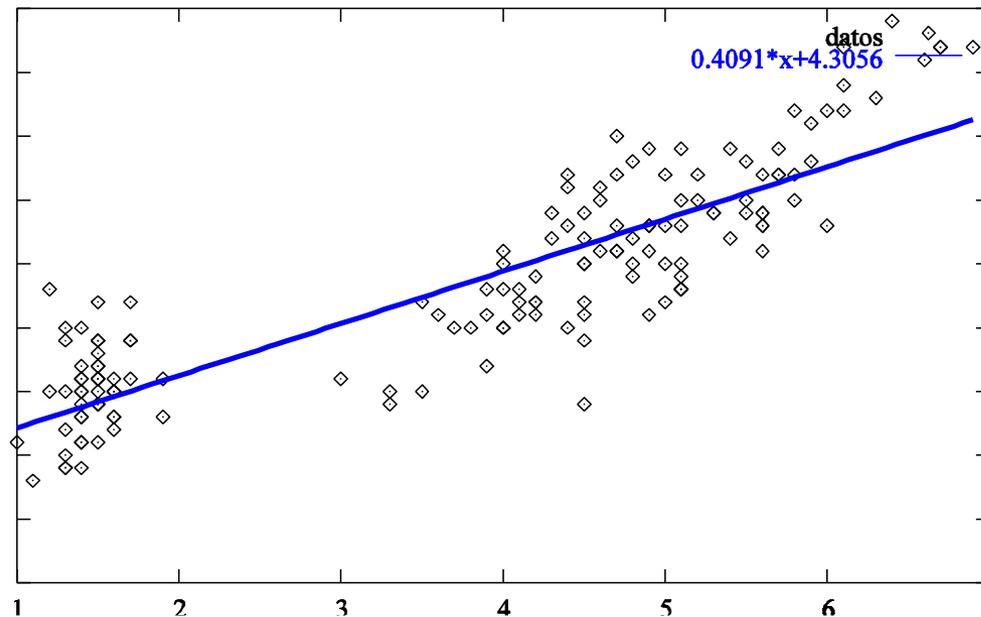
- Son **árboles de regresión** en los que la poda se realiza en mayor medida y en las **hojas** en lugar de un valor numérico **contienen una ecuación de regresión local** a esa partición del espacio



# 6. Árboles de modelos

---

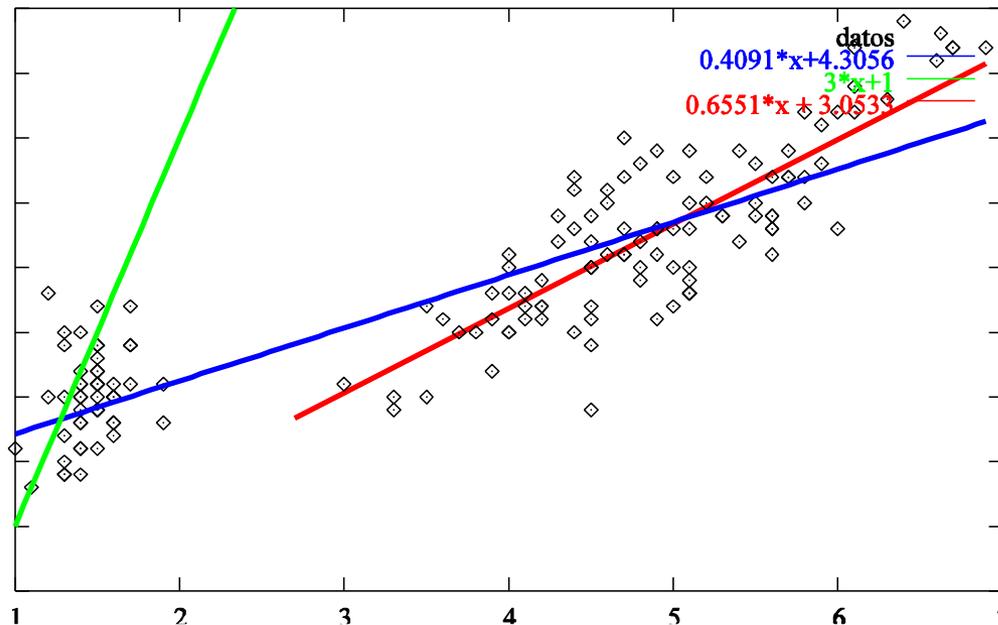
- Son árboles de regresión en los que la poda se realiza en mayor medida y en las hojas en lugar de un valor numérico contienen una ecuación de regresión local a esa partición del espacio



# 6. Árboles de modelos

---

- Son árboles de regresión en los que la poda se realiza en mayor medida y en las hojas en lugar de un valor numérico contienen una ecuación de regresión local a esa partición del espacio



# 6. Árboles de modelos

---

MMAX  $\leq$  14000 :

| CACH  $\leq$  8.5 : LM1 (108/3.99%)

| CACH  $>$  8.5 : LM2 (33/3.89%)

MMAX  $>$  14000 :

| MMAX  $\leq$  22500 : LM3 (37/4.73%)

| MMAX  $>$  22500 : LM4 (31/69.2%)

LM1: class = 15.9 - 0.00453MYCT + 0.00327MMAX

LM2: class = -0.609 + 0.004MMAX + 0.59CACH + 1.57CHMIN

LM3: class = 1.64 - 0.0266MYCT + 0.00485MMIN + 0.00346MMAX +  
0.627CACH + 1.43CHMIN + 0.127CHMAX

LM4: class = -350 - 0.843MYCT + 0.0183MMAX + 1.62CACH

EMR 10.0402

EAR 11.4899 %

WEKA: sin suavizar

poda=0

# Inteligencia de Negocio

## TEMA 3. Modelos de Predicción: Clasificación, regresión y series temporales

---

### Regresión (Predicción numérica)

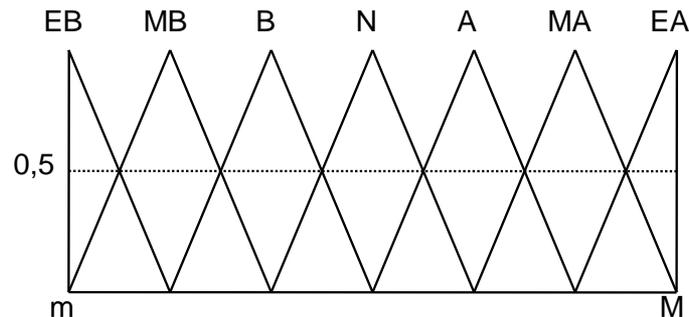
1. Definición del problema
2. Validación de algoritmos de regresión
3. Técnicas de clasificación válidas para regresión
4. Análisis de regresión
5. Árboles de regresión
6. Árboles de modelos
7. Sistemas basados en reglas difusas

# 7. Sistemas basados en reglas difusas

---

- El modelado se puede realizar con **Sistemas Basados en Reglas Difusas (SBRDs)**, que contienen reglas del tipo:

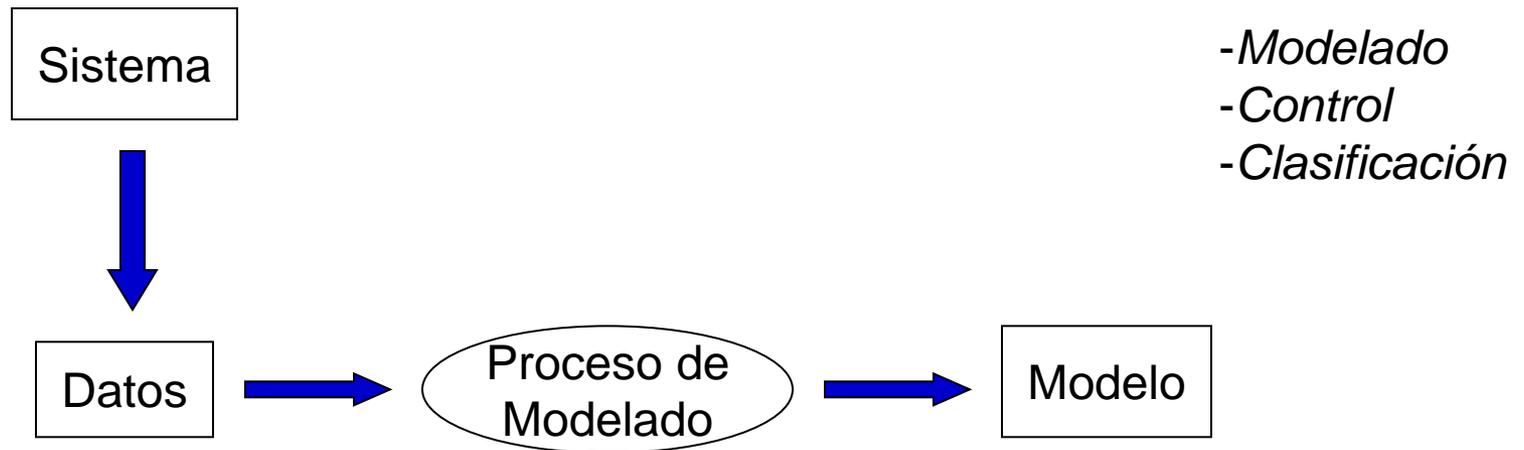
**Si** presión\_atmosférica **es** *Baja* **ENTONCES** probabilidad\_lluvia **es** *Alta*



- Existen distintas clases de modelado con SBRDs:
  - Modelado Difuso Lingüístico:** Atiende al poder descriptivo de los SBRDs
  - Modelado Difuso Preciso:** Atiende al poder aproximativo de los SBRDs

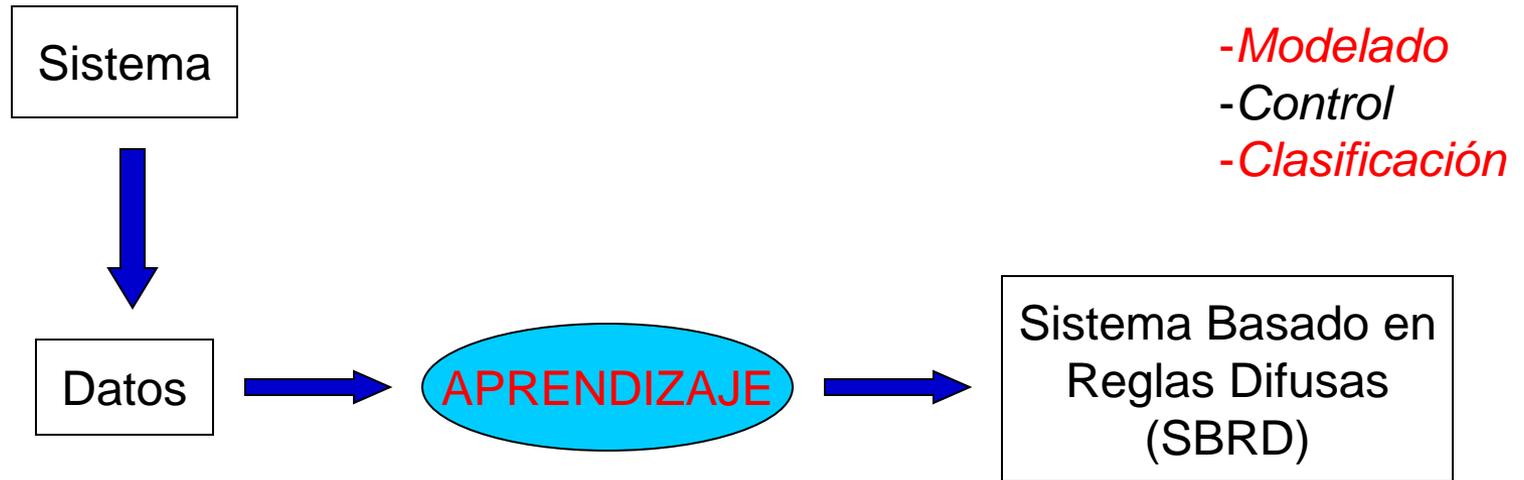
# 7. Sistemas basados en reglas difusas

---



# 7. Sistemas basados en reglas difusas

---

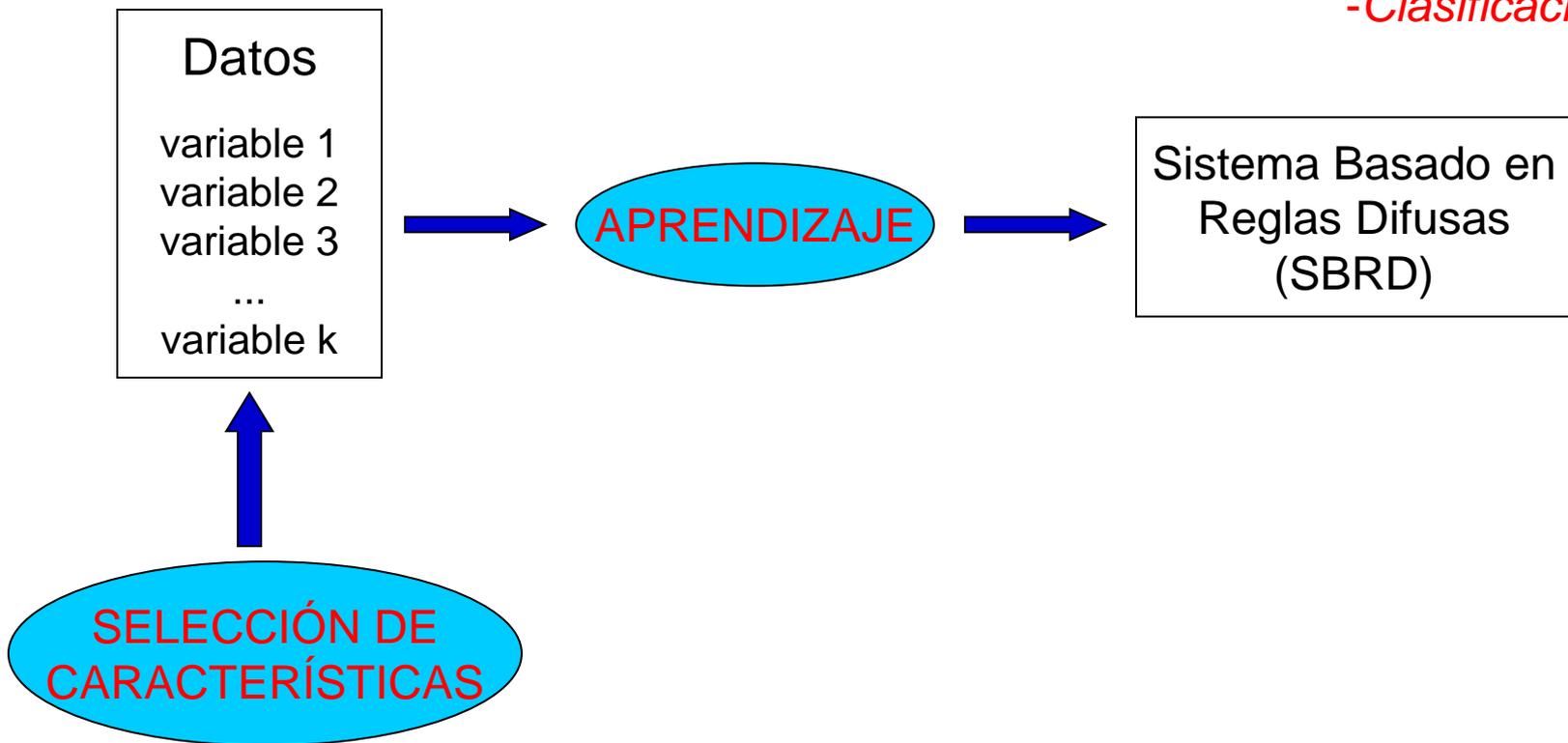


# 7. Sistemas basados en reglas difusas

---

Variables redundantes  
o irrelevantes

*-Modelado*  
*-Control*  
*-Clasificación*

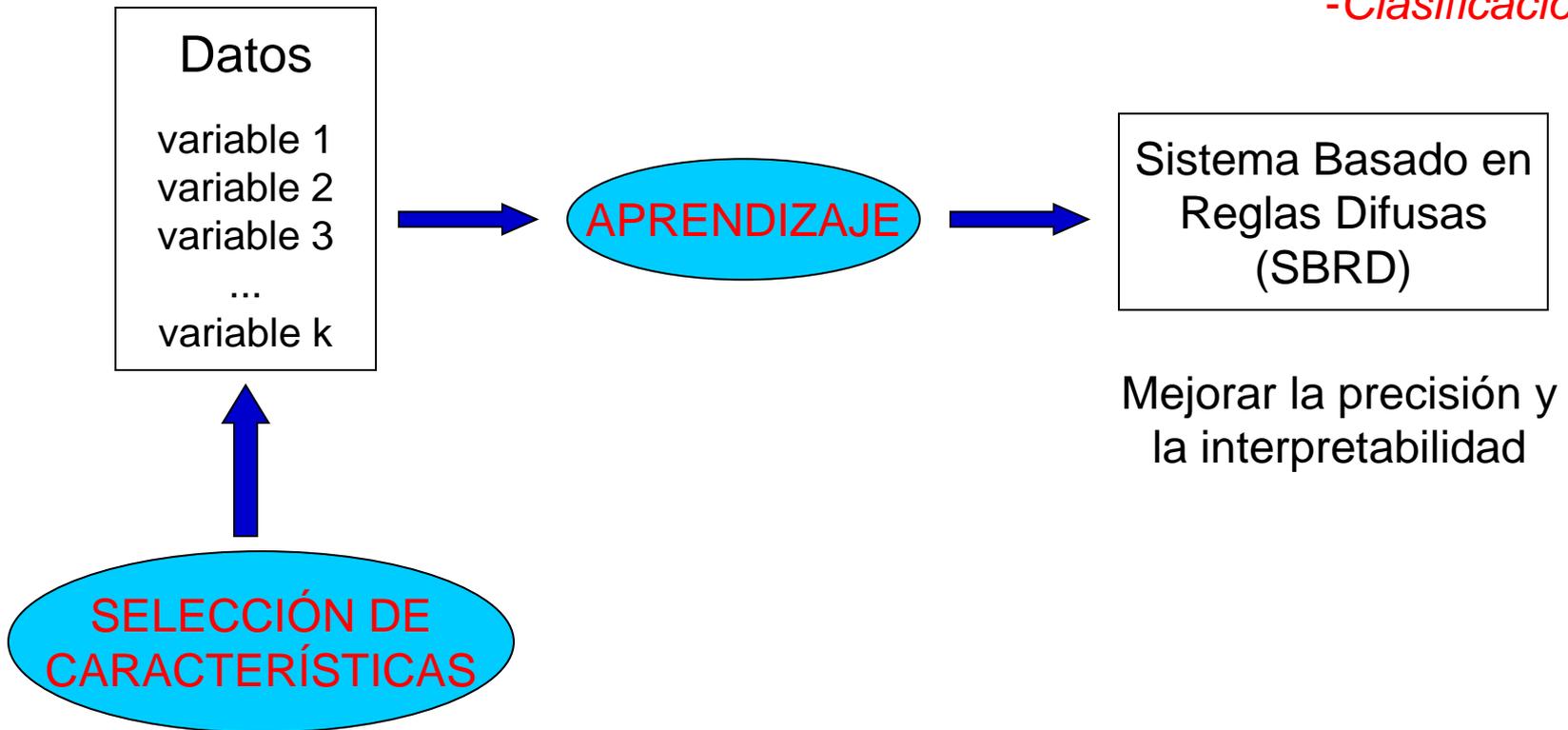


# 7. Sistemas basados en reglas difusas

---

Variables redundantes  
o irrelevantes

*-Modelado*  
*-Control*  
*-Clasificación*



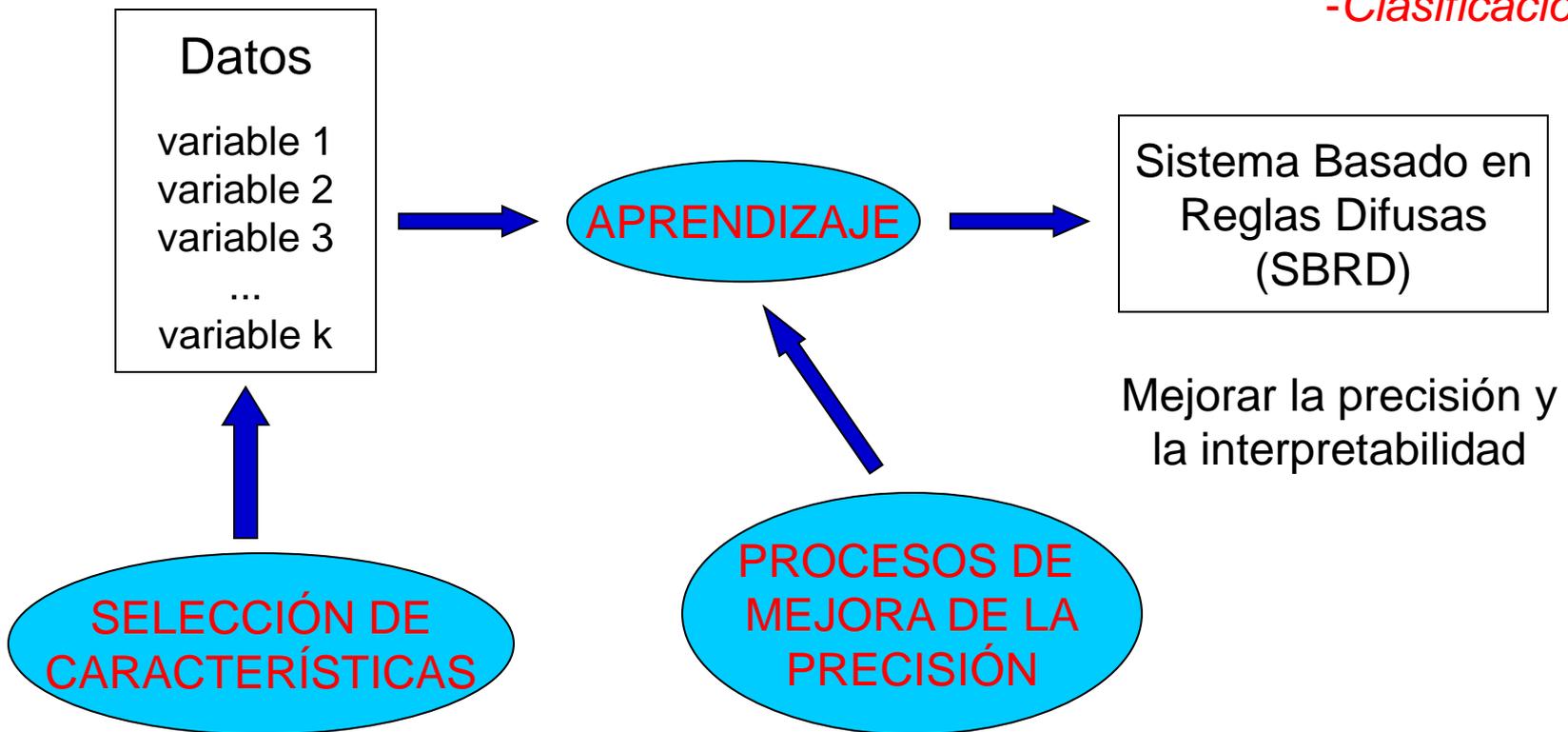
Mejorar la precisión y  
la interpretabilidad

# 7. Sistemas basados en reglas difusas

---

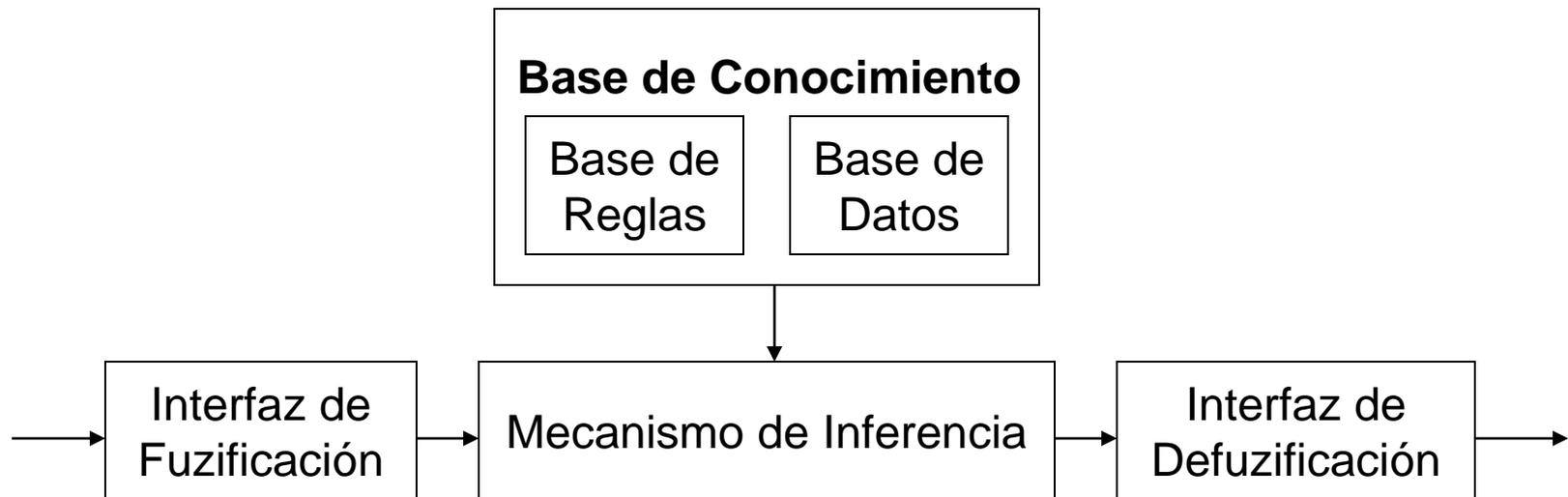
Variables redundantes  
o irrelevantes

*-Modelado*  
*-Control*  
*-Clasificación*



# 7. Sistemas basados en reglas difusas

---

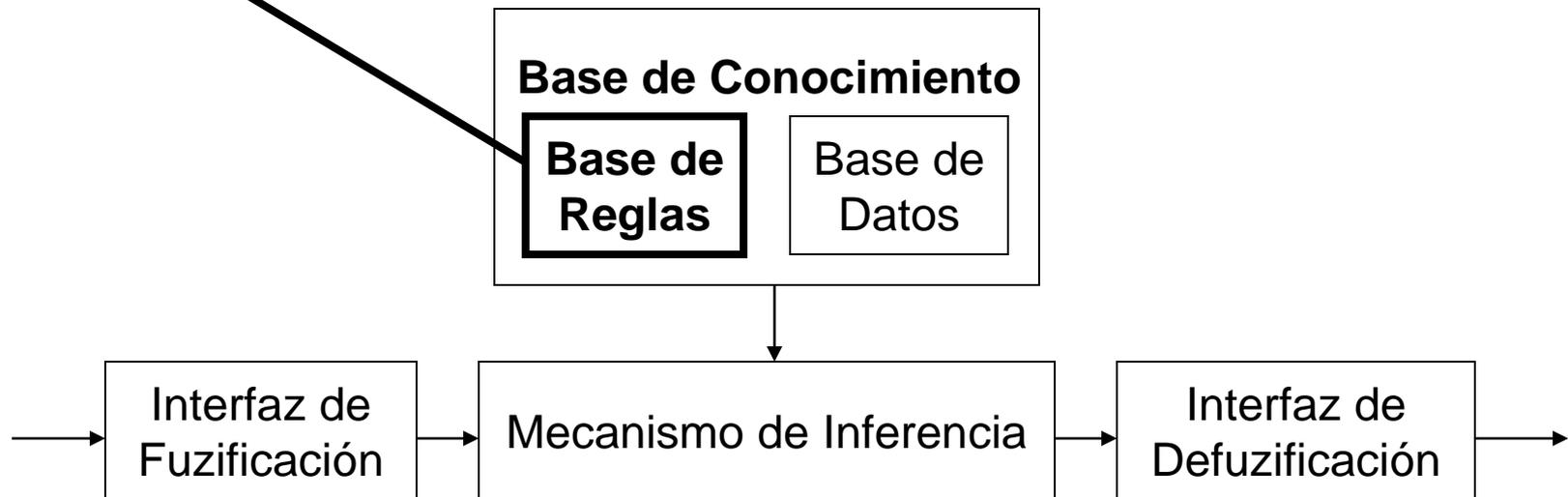


## **Sistema Basado en Reglas Difusas**

# 7. Sistemas basados en reglas difusas

---

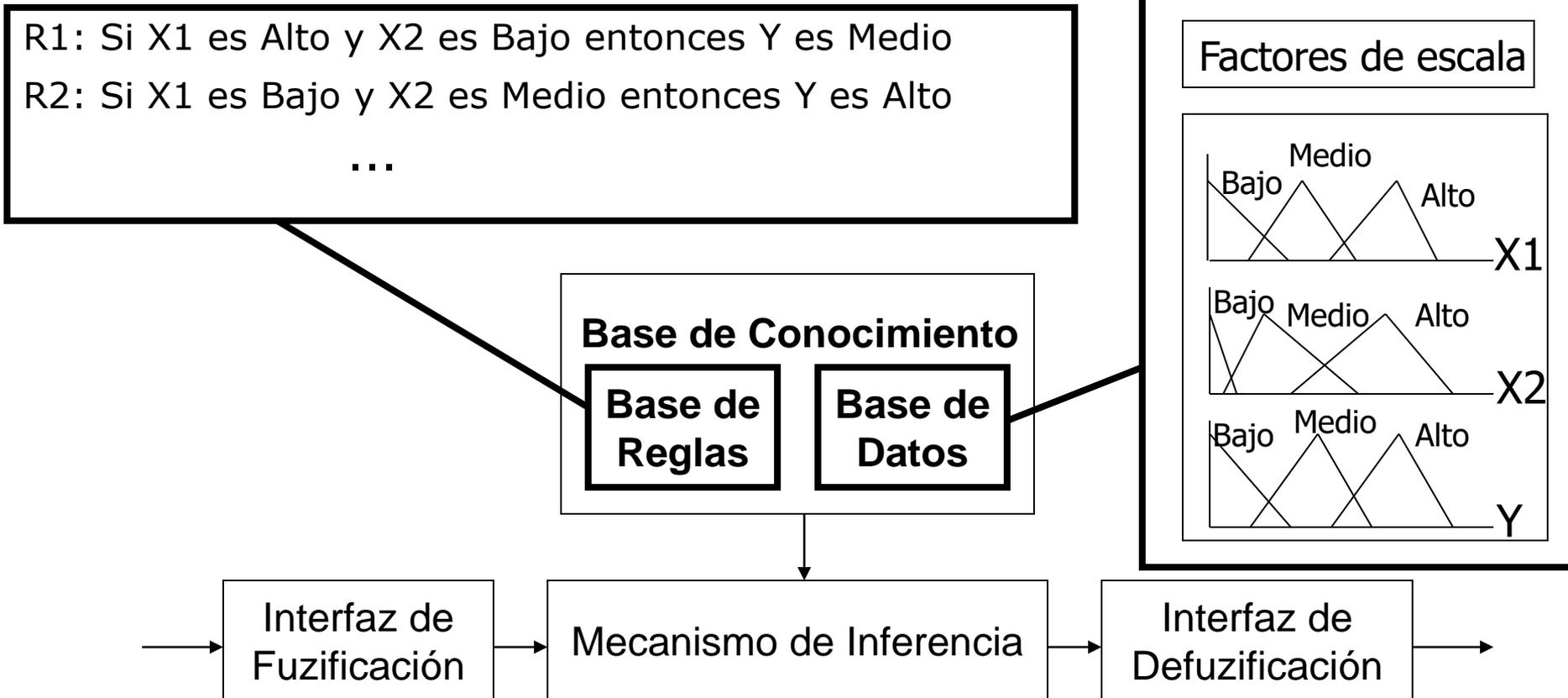
R1: Si X1 es Alto y X2 es Bajo entonces Y es Medio  
R2: Si X1 es Bajo y X2 es Medio entonces Y es Alto  
...



**Sistema Basado en Reglas Difusas**

# 7. Sistemas basados en reglas difusas

---



**Sistema Basado en Reglas Difusas**

# 7. Sistemas basados en reglas difusas

---

R1: Si X1 es Alto y X2 es Bajo entonces Y es Medio  
R2: Si X1 es Bajo y X2 es Medio entonces Y es Alto  
...

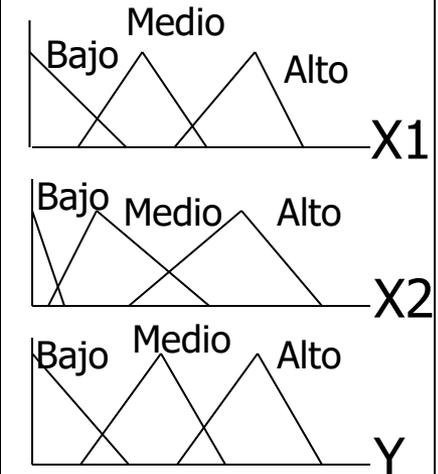
...

**Base de Conocimiento**

**Base de Reglas**

**Base de Datos**

**Factores de escala**



Entrada escalada

Interfaz de Fuzificación

Mecanismo de Inferencia

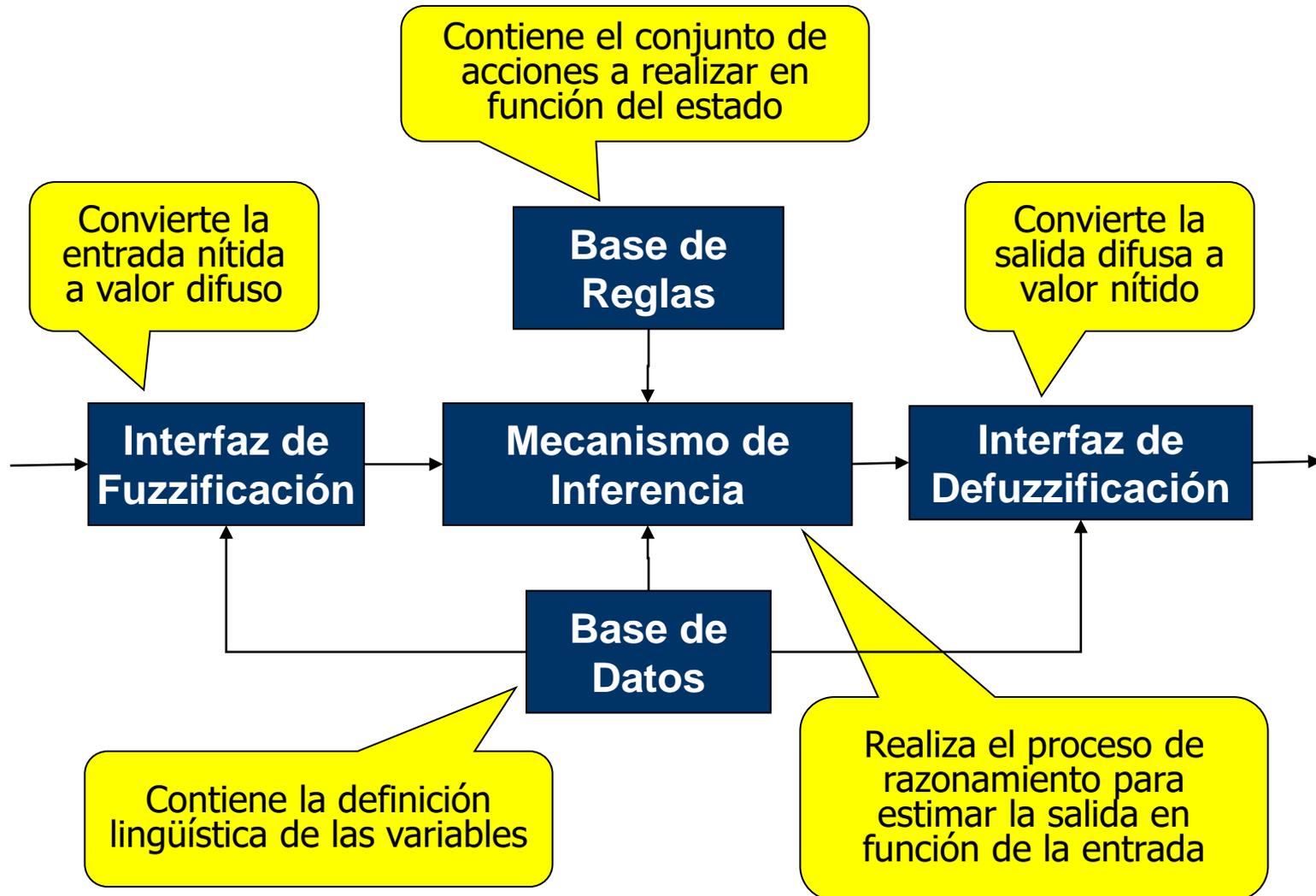
Interfaz de Defuzificación

Salida escalada

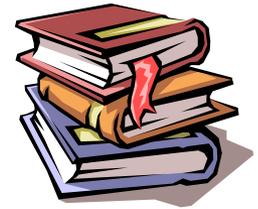
**Sistema Basado en Reglas Difusas**

# 7. Sistemas basados en reglas difusas

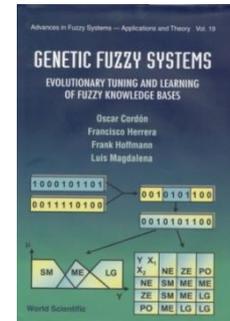
---



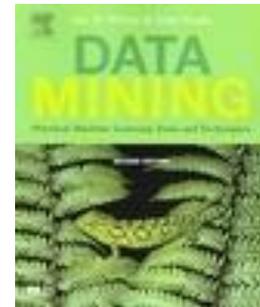
# Bibliografía



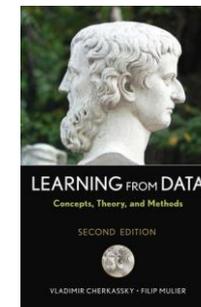
O. Cordón, F. Herrera, F. Hoffmann, L. Magdalena  
GENETIC FUZZY SYSTEMS.  
Evolutionary Tuning and Learning of  
Fuzzy Knowledge Bases.  
World Scientific, Julio 2001.



I.H. Witten, E. Frank.  
Data Mining: Practical Machine Learning Tools and Techniques,  
Second Edition, Morgan Kaufmann, 2005.  
<http://www.cs.waikato.ac.nz/~ml/weka/book.html>



V. Cherkassky, F.M. Mulier  
Learning from Data: Concepts, Theory, and Methods,  
2<sup>nd</sup> Edition, Wiley-IEE Prees, 2007



# Inteligencia de Negocio

## TEMA 3. Modelos de Predicción: Clasificación, regresión y series temporales

---

1. Clasificación
2. Regresión
- 3. Series Temporales**

# Inteligencia de Negocio

## TEMA 4. Modelos de Predicción: Clasificación, regresión y series temporales

---

1. Clasificación
2. Regresión
- 3. Series Temporales**

### **Bibliografía**

R. Hyndman, G. Athanasopoulos, «Forecasting and time series» 2013  
(Disponible en <https://www.otexts.org/fpp>)

R.H. Shumway, D.S. Stoffer, «Time Series Analysis and Its Applications»,  
Springer, 3rd Ed., 2011

# Contents

---

- Forecasting
- Forecaster's toolbox
- Simple regression
- Multivariate regression
- Time series decomposition
- ARIMA models
- Advanced forecasting models

**Agradecimientos: José Manuel Benítez, autor de las transparencias, y que ha cedido para su uso como Tema 3, parte III.**

# Contents

---

- **Forecasting**
- Forecaster's toolbox
- Simple regression
- Multivariate regression
- Time series decomposition
- ARIMA models
- Advanced forecasting models

# Forecasting

---

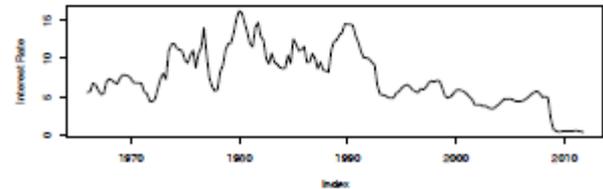
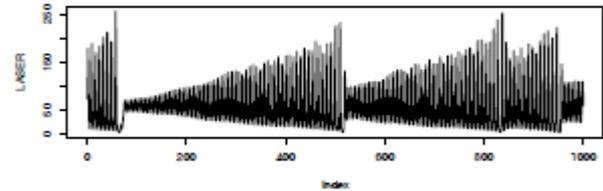
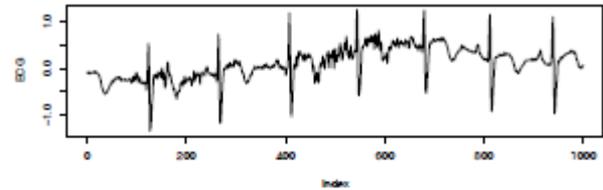
## Definition

- **Forecasting**: Predicting the future as accurately as possible, given all the information available including **historical data** and knowledge of any future events that might impact the forecasts

# Time series

- Anything that is observed over time is a time series
- Time series observed at regular intervals of time (every minute, hourly, daily, weekly, ...)

$$\{X_{t_1}, X_{t_2}, X_{t_3}, \dots, X_{t_n}\}$$



# Examples (Time series forecasting)

---

- Forecast of electricity demand: highly accurate
- Forecast on currency exchange rates: rough



**Time series forecasting:** Time series data is useful when you are forecasting something that is changing over time (e.g., stock prices, sales, profits, ...)

# Quantitative forecasting

---

- Can be applied when:
  - **Numerical** data about the past is **available**
  - It is reasonable to assume that **some aspects** of the **past** patterns will **continue** into the **future**

# Factors affecting forecast

---

- Time horizon

Time Frame (How far can we predict?)

short-term (1 - 2 periods)

medium-term (5 - 10 periods)

long-term (12+ periods)

- Types of data patterns

# Time series forecasting

---

- Time series data is useful when you are forecasting something that is changing over time (e.g., stock prices, sales, profits, ...)
- Time series forecasting intends to estimate how the sequence of observations will continue in the future

Time Frame (How far can we predict?)

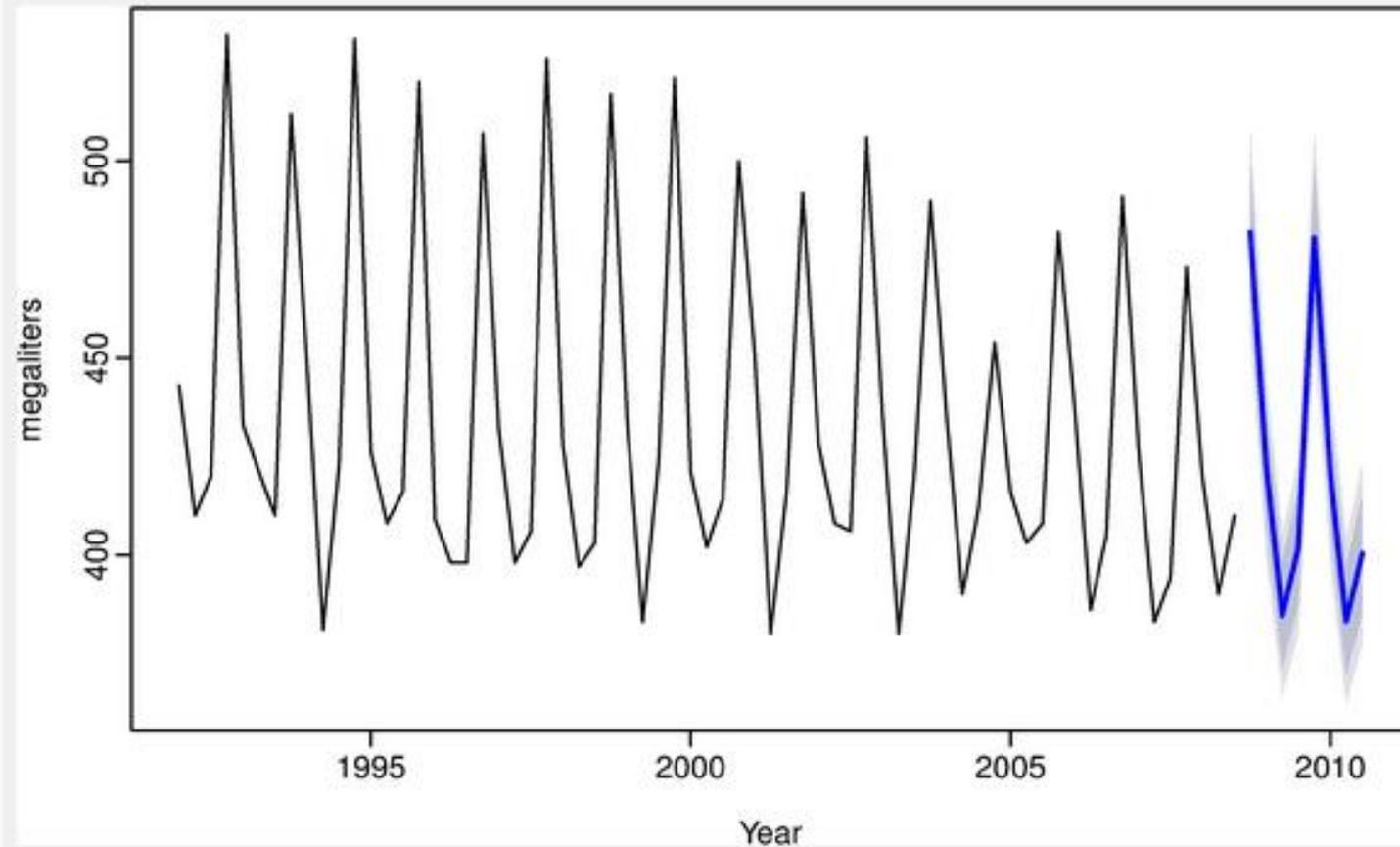
short-term (1 - 2 periods)

medium-term (5 - 10 periods)

long-term (12+ periods)

# Beer production forecast

---



# Contents

---

- Forecasting
- **Forecaster's toolbox**
- Simple regression
- Multivariate regression
- Time series decomposition
- ARIMA models
- Advanced forecasting models

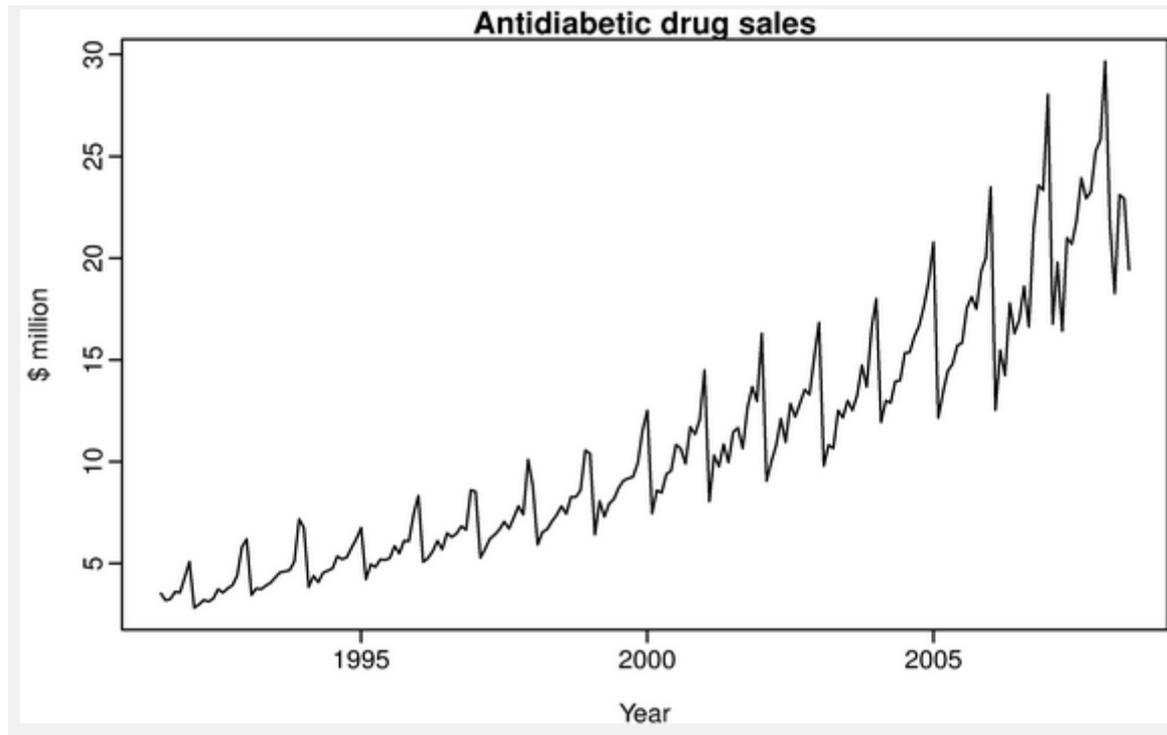
# Time plot

---



## Time plot, example 2

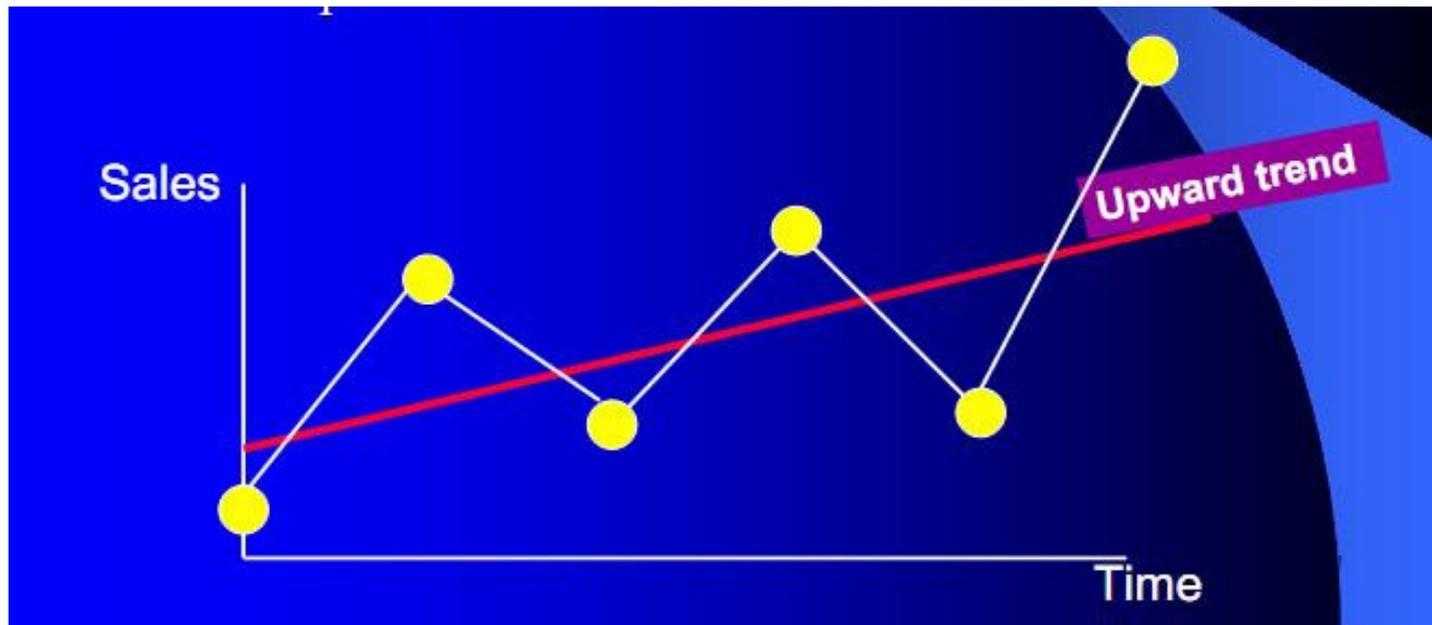
---



# Time Series patterns

---

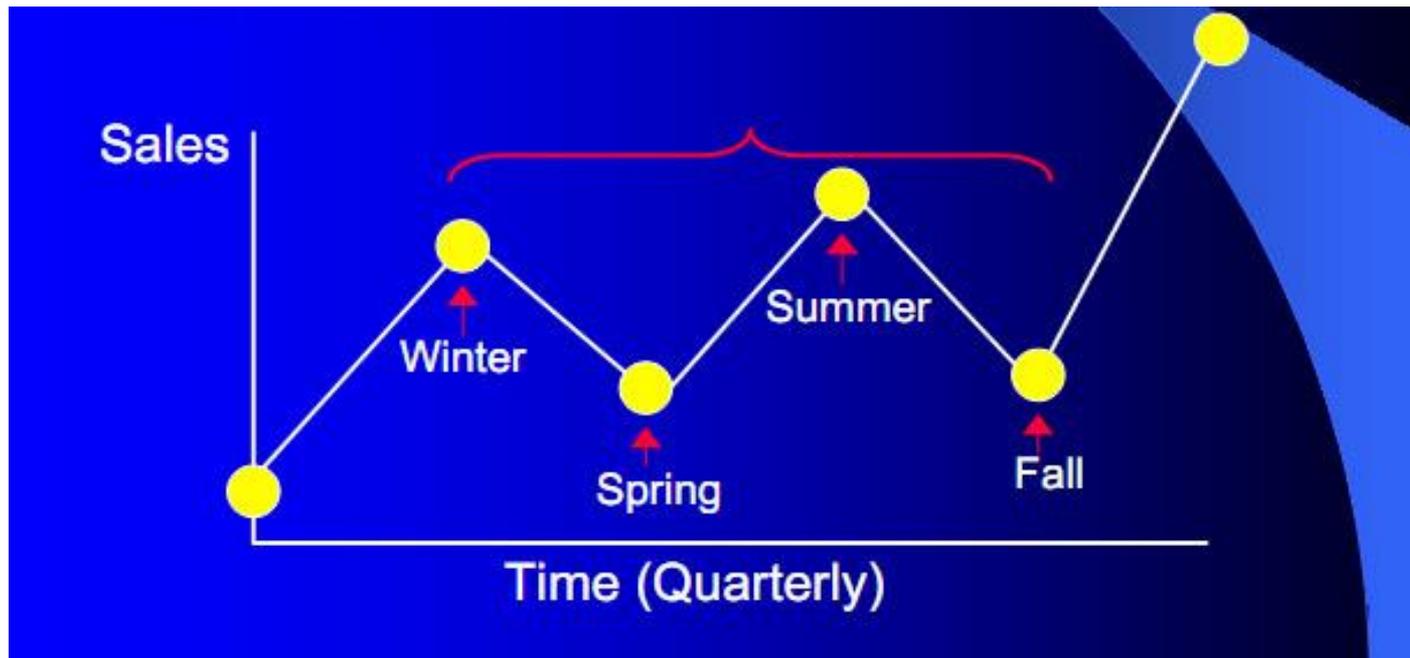
- *Trend*: long-term increase or decrease in the data



# Time Series patterns

---

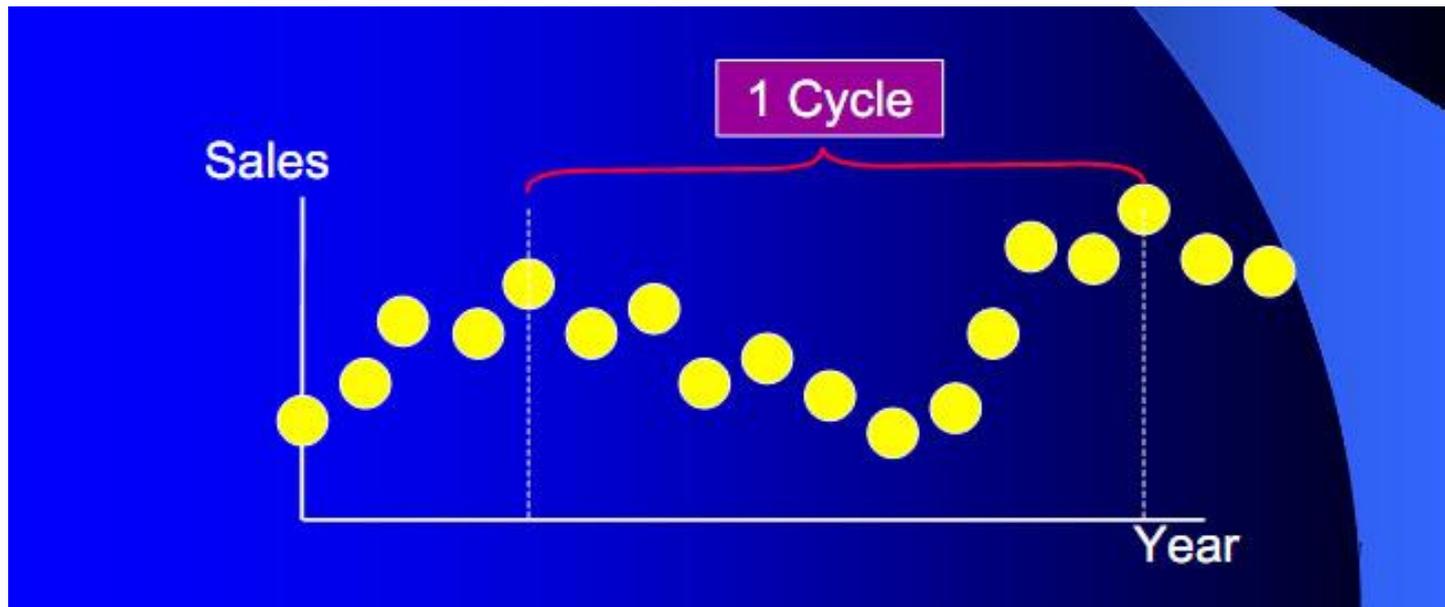
- *Seasonal pattern*: data affected by seasonal factors such as time of the year or day of the week



# Time Series patterns

---

- *Cycle*: data exhibits rises and falls that are not of a fixed period; variable and unknown length



# Numerical data summaries

---

## ■ Univariate statistics

- Average
- Median
- Percentiles
- Interquartile Range (IQR)
- Standard deviation

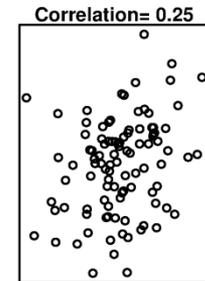
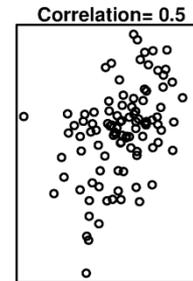
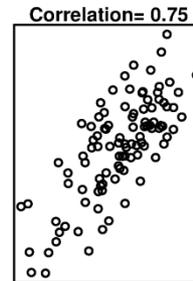
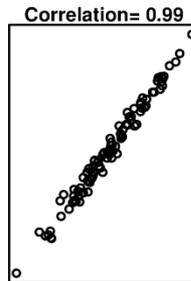
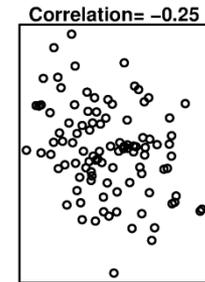
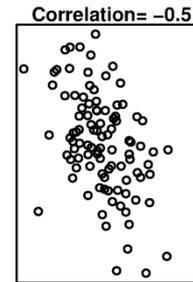
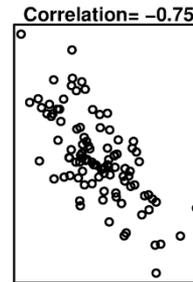
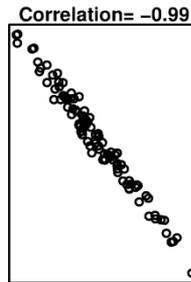
$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = (x_1 + x_2 + x_3 + \cdots + x_N)/N.$$

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}.$$

# Correlation coefficient

Strength of linear relationship

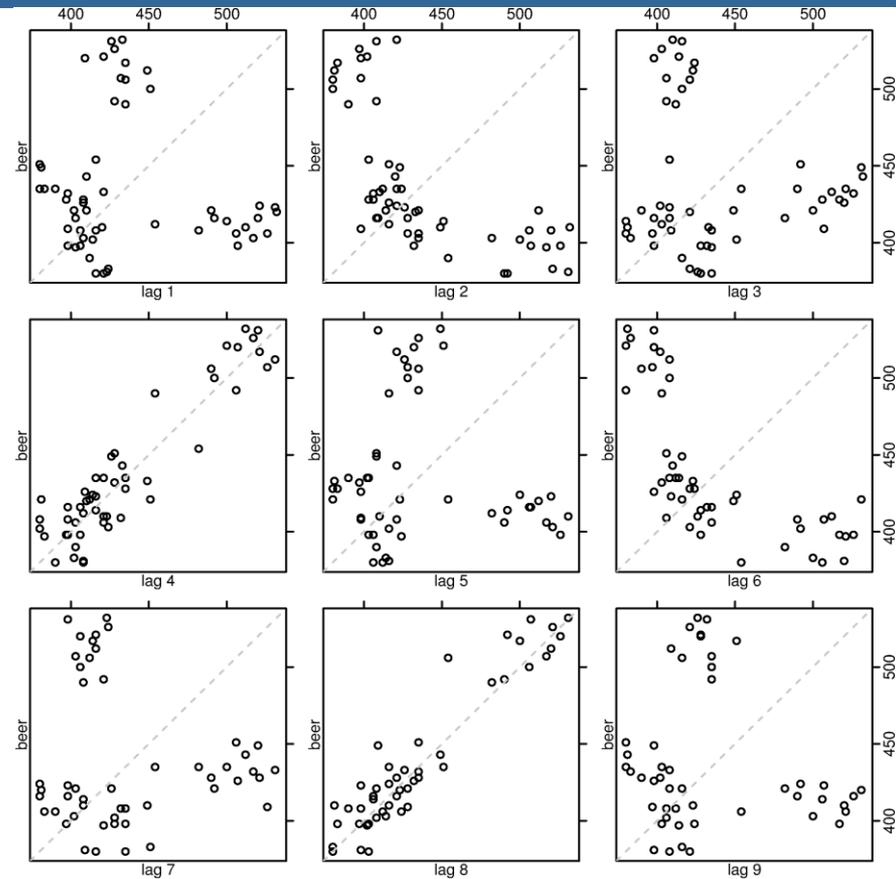
$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}}$$



# Autocorrelation

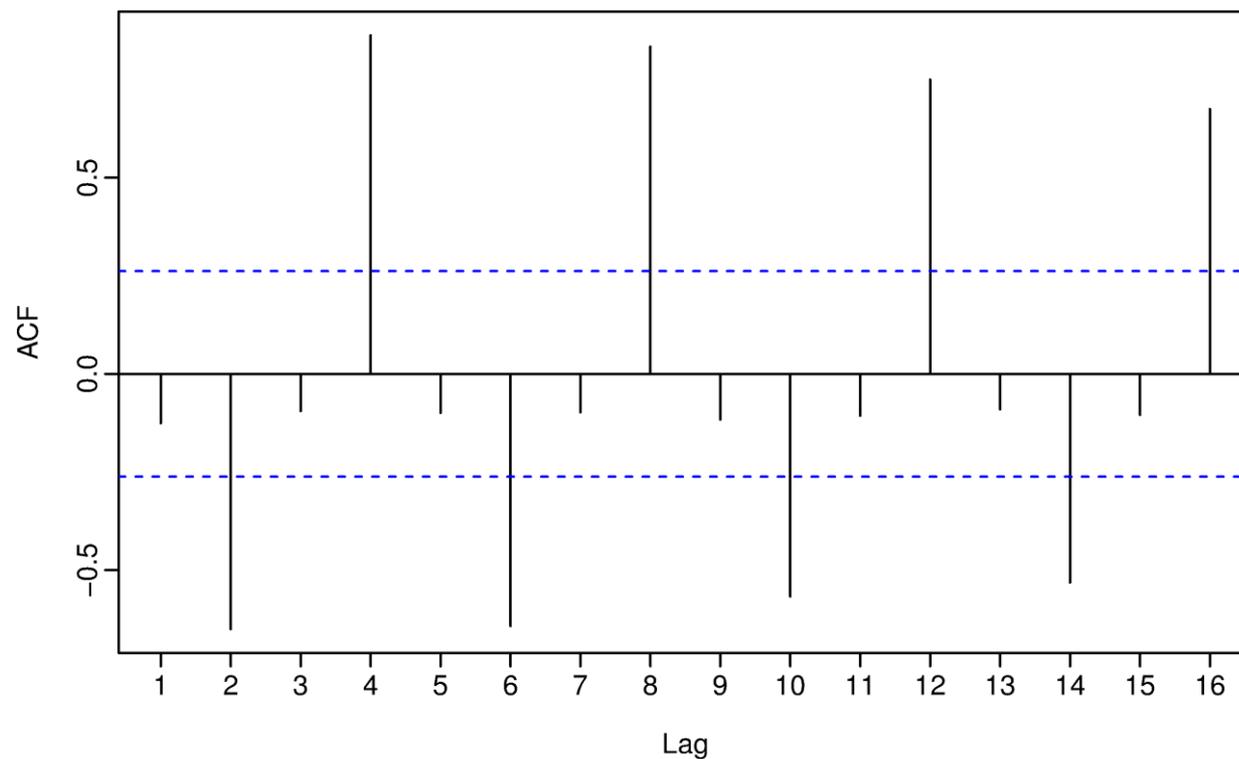
- Relationship between lagged values of a time series

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2},$$



# Autocorrelation function (ACF)

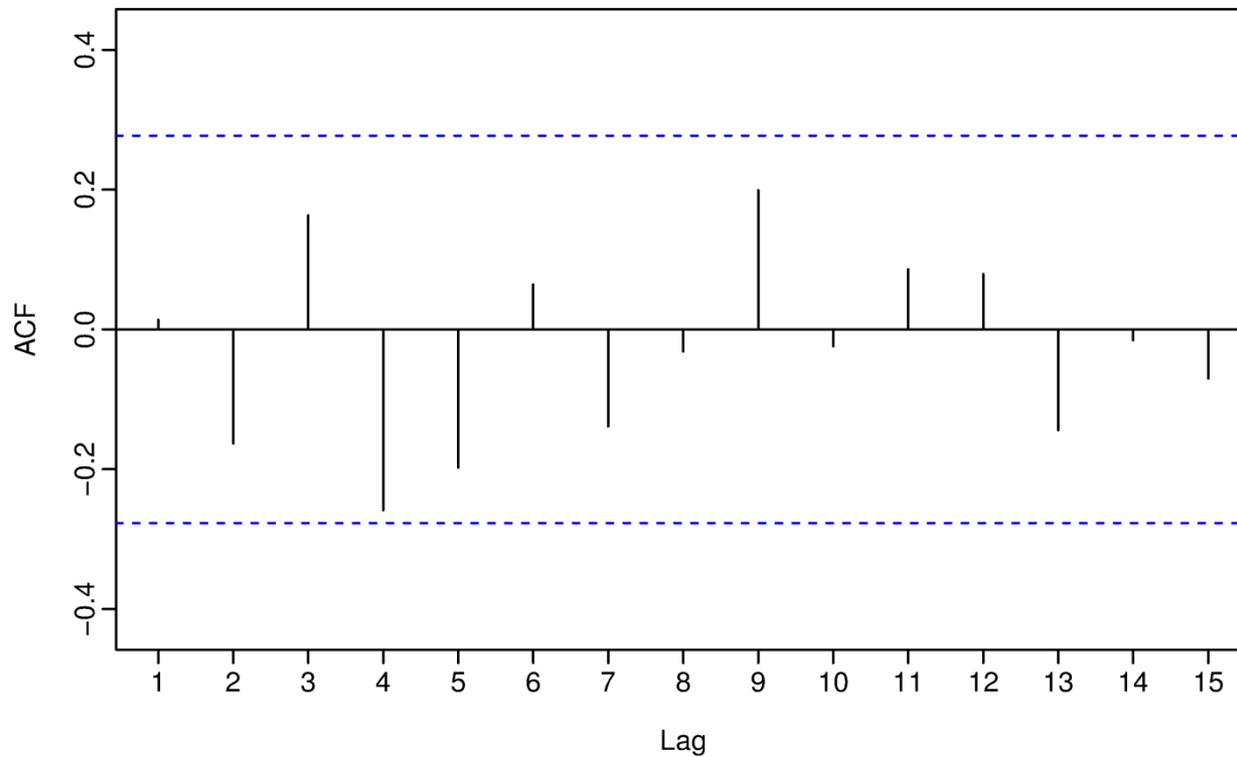
$r_1$	$r_2$	$r_3$	$r_4$	$r_5$	$r_6$	$r_7$	$r_8$	$r_9$
-0.126	-0.650	-0.094	0.863	-0.099	-0.642	-0.098	0.834	-0.116



# White noise

---

- Series showing NO autocorrelation



# Simple forecasting methods

---

- Average method

$$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \dots + y_T)/T.$$

- Naïve method

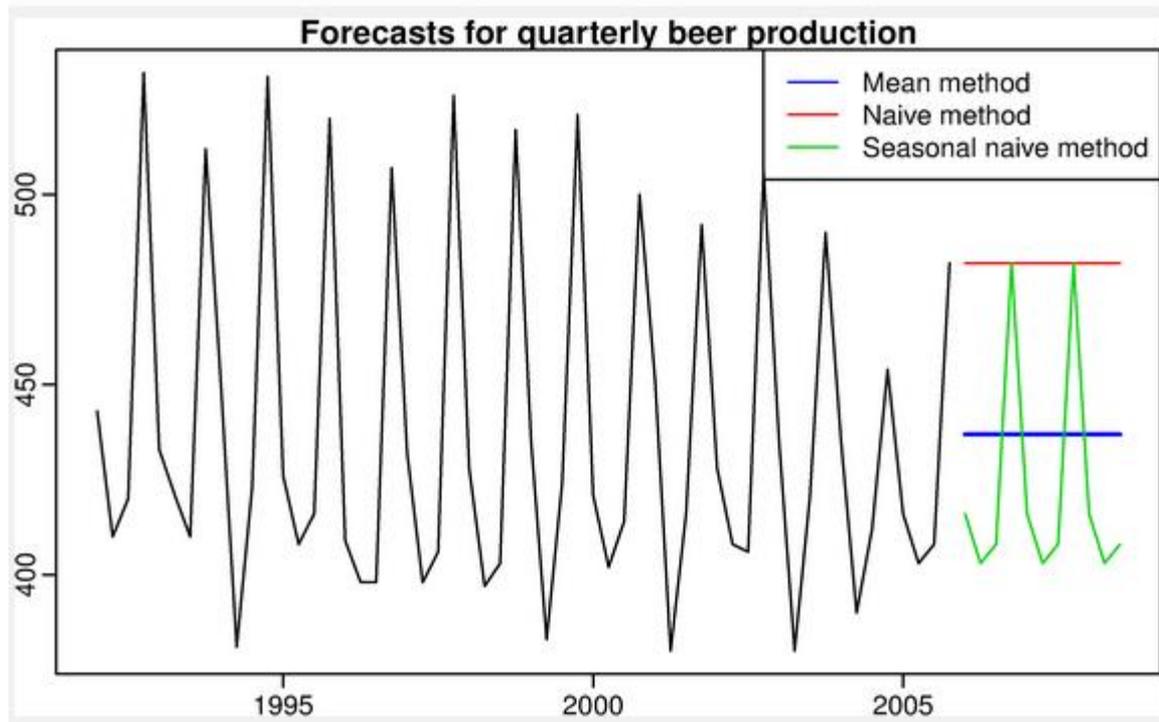
- Forecast: last value

- Seasonal naïve method

- Drift method

- The forecast increases or decreases over time, by the average in historical data

# Simple forecasting methods (2)



# Transformations

---

- Adjusting historical data can lead to a simpler forecasting model
- Mathematical transformation
- Calendar adjustments
- Population adjustments
- Inflation adjustments

# Evaluating forecast accuracy

---

- Forecast error:

$$e_i = y_i - \hat{y}_i$$

- Scale-dependent errors

$$\text{MAE} = \text{mean}(|e_i|),$$

$$\text{RMSE} = \sqrt{\text{mean}(e_i^2)}.$$

$$p_i = 100e_i/y_i$$

- Percentage error:

- Scaled errors

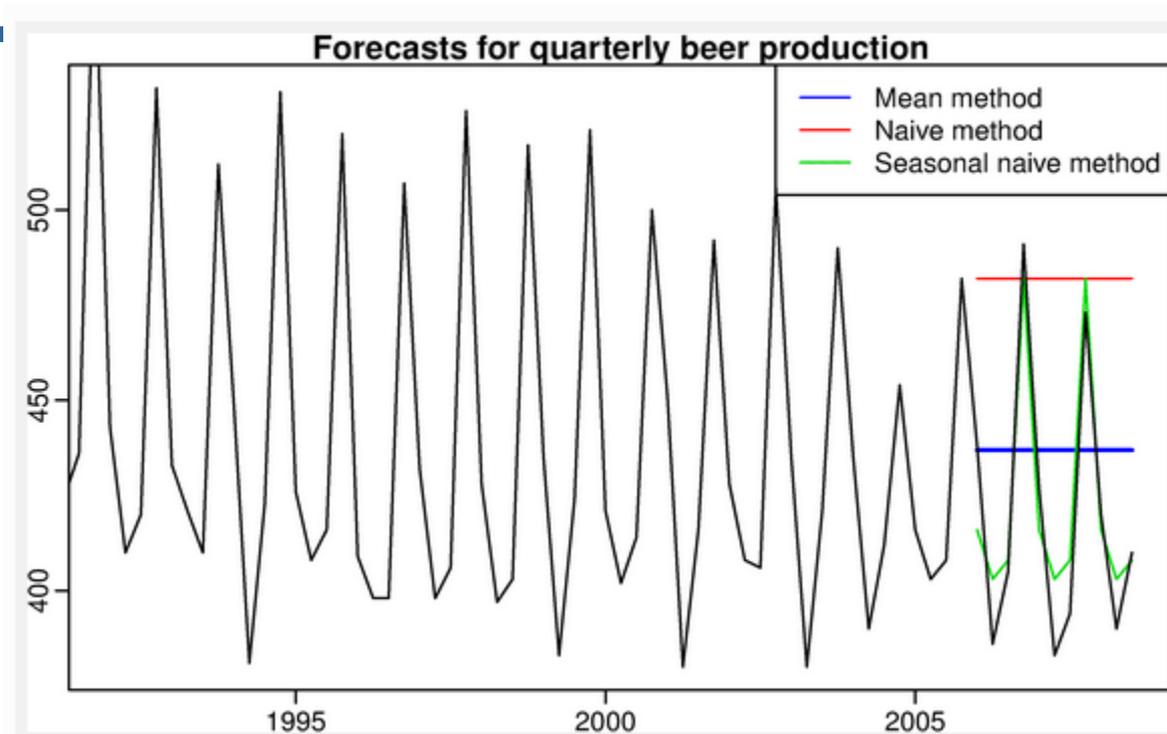
$$\text{MAPE} = \text{mean}(|p_i|).$$

$$\text{sMAPE} = \text{mean}(200|y_i - \hat{y}_i|/(y_i + \hat{y}_i))$$

$$q_j = \frac{e_j}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|}.$$

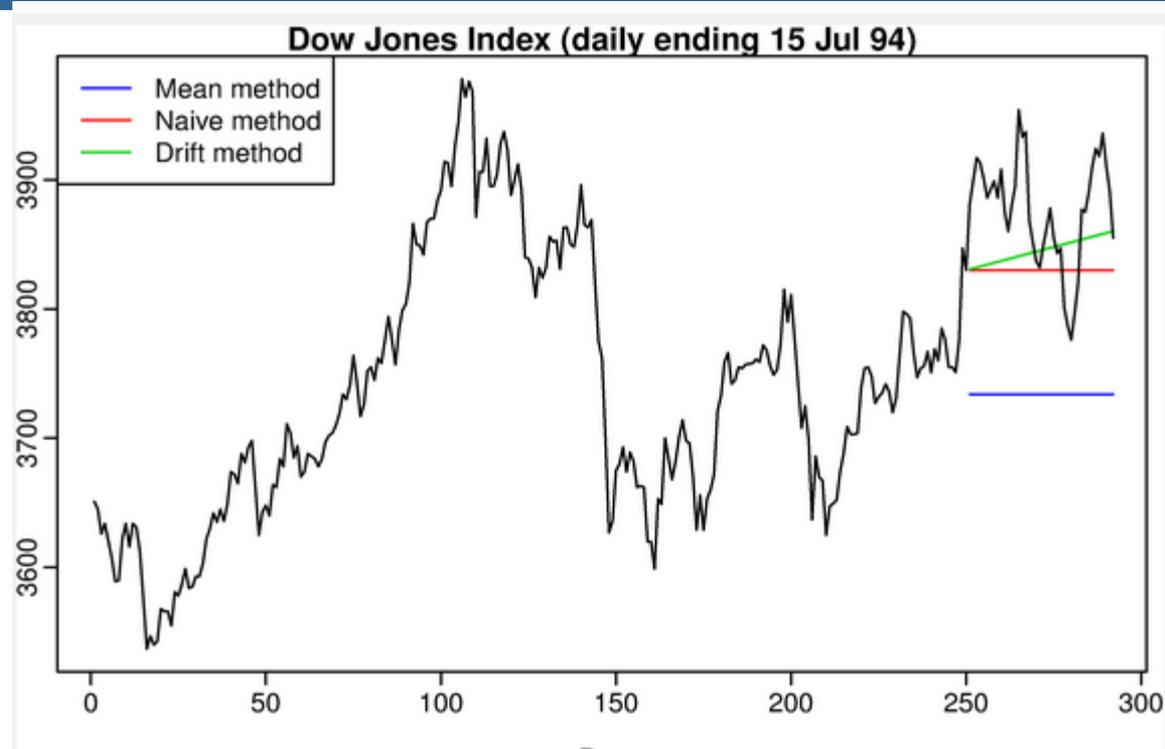
$$\text{MASE} = \text{mean}(|q_j|).$$

# Seasonal time series



Method	RMSE	MAE	MAPE	MASE
Mean method	38.01	33.78	8.17	0.61
Naïve method	70.91	63.91	15.88	1.15
Seasonal naïve method	12.97	11.27	2.73	0.20

# Non-seasonal time series



Method	RMSE	MAE	MAPE	MASE
Mean method	148.24	142.42	3.66	8.70
Naïve method	62.03	54.44	1.40	3.32
Drift method	53.70	45.73	1.18	2.79

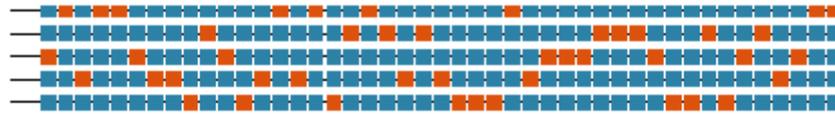
# Methodology

---

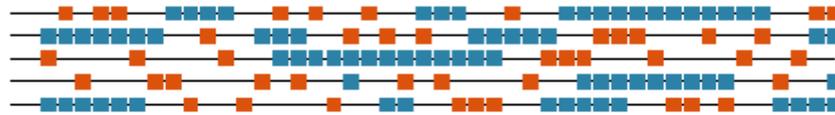
- As in any other modeling task it is essential to conduct a right evaluation
- Data should be split into training and test parts
- Improved through Cross-validation
- Even further improved through Blocked Cross-Validation

# Model selection procedures

---



cross-validation



non-dep. cross-validation



blocked cross-validation



last block

# Residual diagnostics

---

- Residual:  $e_i = y_i - \hat{y}_i$
- *Good* forecasting method:
  - Residuals uncorrelated
  - Residuals have zero mean
- If the method does not fulfill them, it can be improved
- Additional properties:
  - Residuals have constant variance
  - Residuals are normally distributed

# Contents

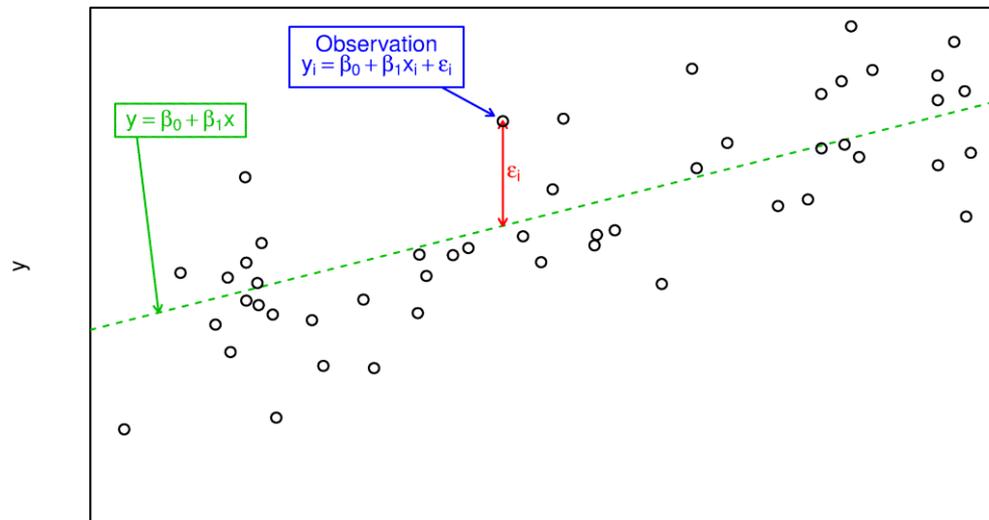
---

- Forecasting
- Forecaster's toolbox
- **Simple regression**
- Multivariate regression
- Time series decomposition
- ARIMA models
- Advanced forecasting models

# Simple linear model

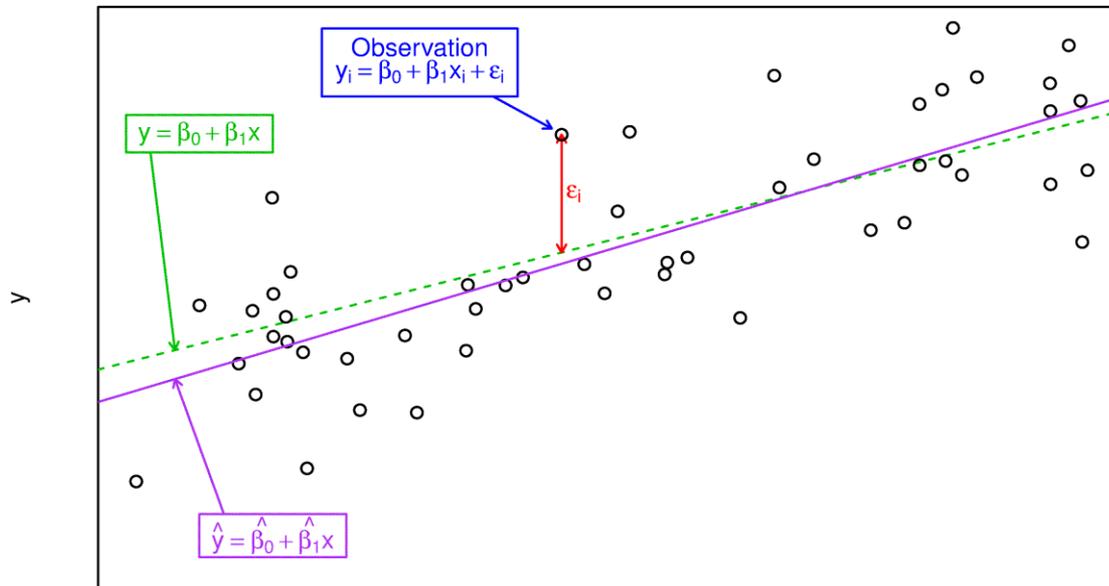
- Let's assume that the forecast and predictor variables are linearly related:

$$y = \beta_0 + \beta_1 x + \varepsilon$$



# Least squares estimation

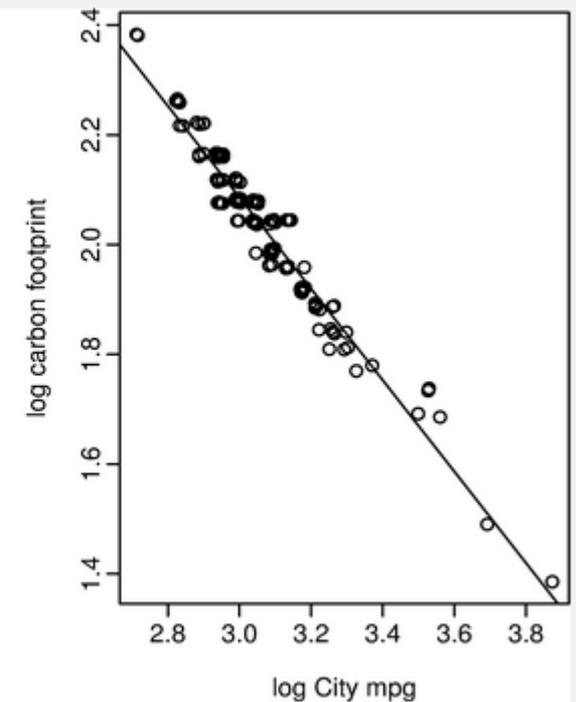
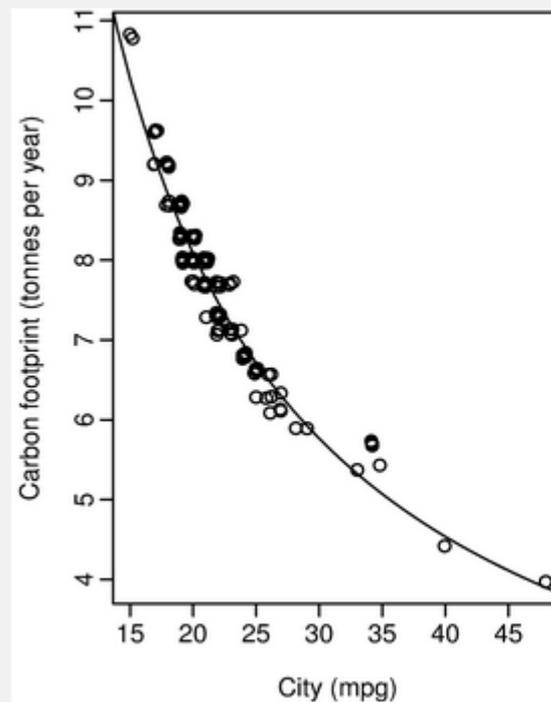
$$\sum_{i=1}^N \epsilon_i^2 = \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2$$



# Non-linear regression

- A non-linear functional form may be more suitable for a problem than a linear one
- This can be obtained through transformation of  $y$  or  $x$

$$\log y_i = \beta_0 + \beta_1 \log x_i + \varepsilon_i.$$



# Contents

---

- Forecasting
- Forecaster's toolbox
- Simple regression
- **Multivariate regression**
- Time series decomposition
- Exponential smoothing
- ARIMA models
- Advanced forecasting models

# Multiple regression

---

One variable to be forecast and **several predictor variables**

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i} + e_i$$

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{1,i} - \cdots - \beta_k x_{k,i})^2$$

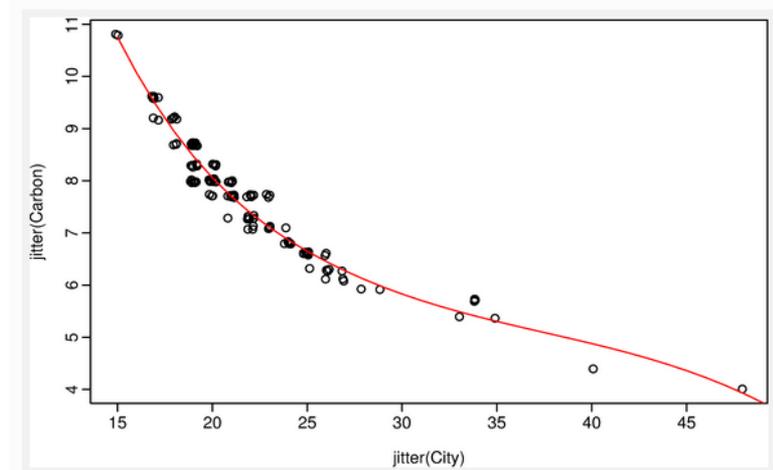
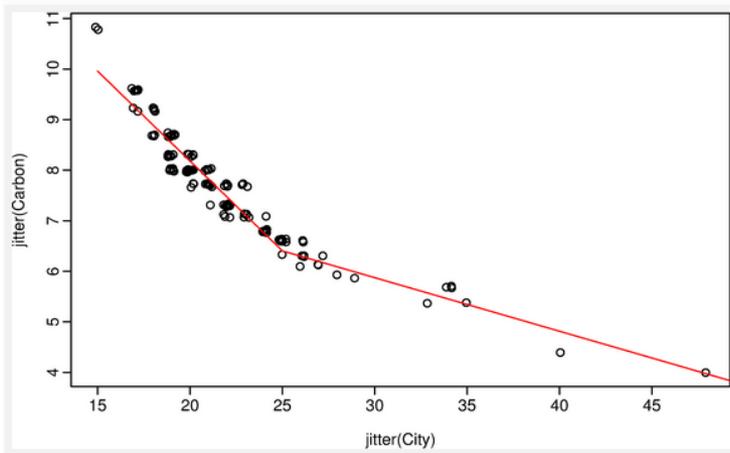
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$$

Predicting Electricity Demand (ED)

*ED = f(current temperature, strength of economy, population, time of day, day of week, error)*

# Non-linear regression

$$y = f(x) + e.$$



# Correlation is not causation

---

- A variable  $x$  may be useful for predicting a variable  $y$ , but that does not mean  $x$  is causing  $y$ .
- Correlations are useful for forecasting, even when there is no causal relationship between the two variables

# Contents

---

- Forecasting
- Forecaster's toolbox
- Simple regression
- Multivariate regression
- **Time series decomposition**
- ARIMA models
- Advanced forecasting models

# Time Series decomposition

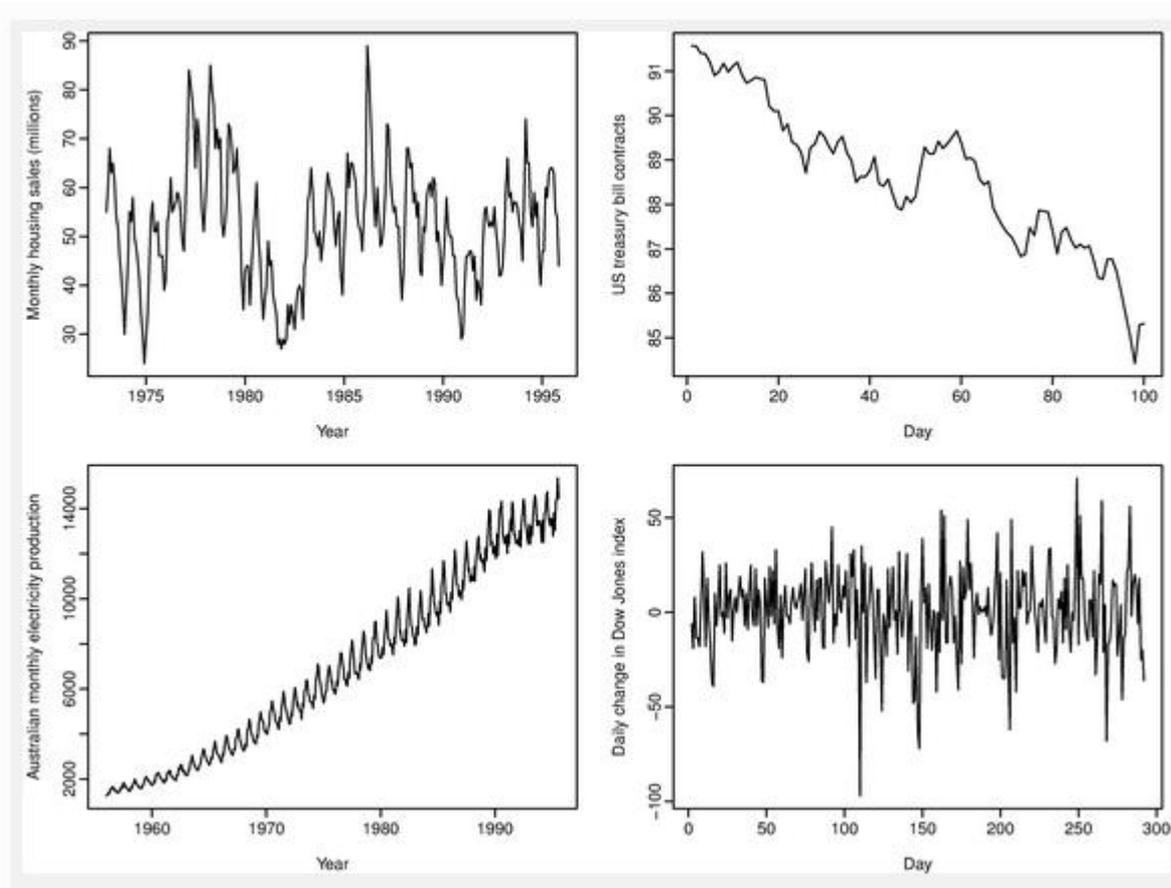
---

- Time series can exhibit a huge variety of patterns and it is helpful to categorize some of the patterns and behaviors that can be seen
- It is also sometimes useful to try to split a time series into several components, each representing one of the underlying components

# Time series components

---

Trend, Seasonal, Cyclic



# Time series decomposition

---

Additive decomposition

$$y_t = S_t + T_t + E_t$$

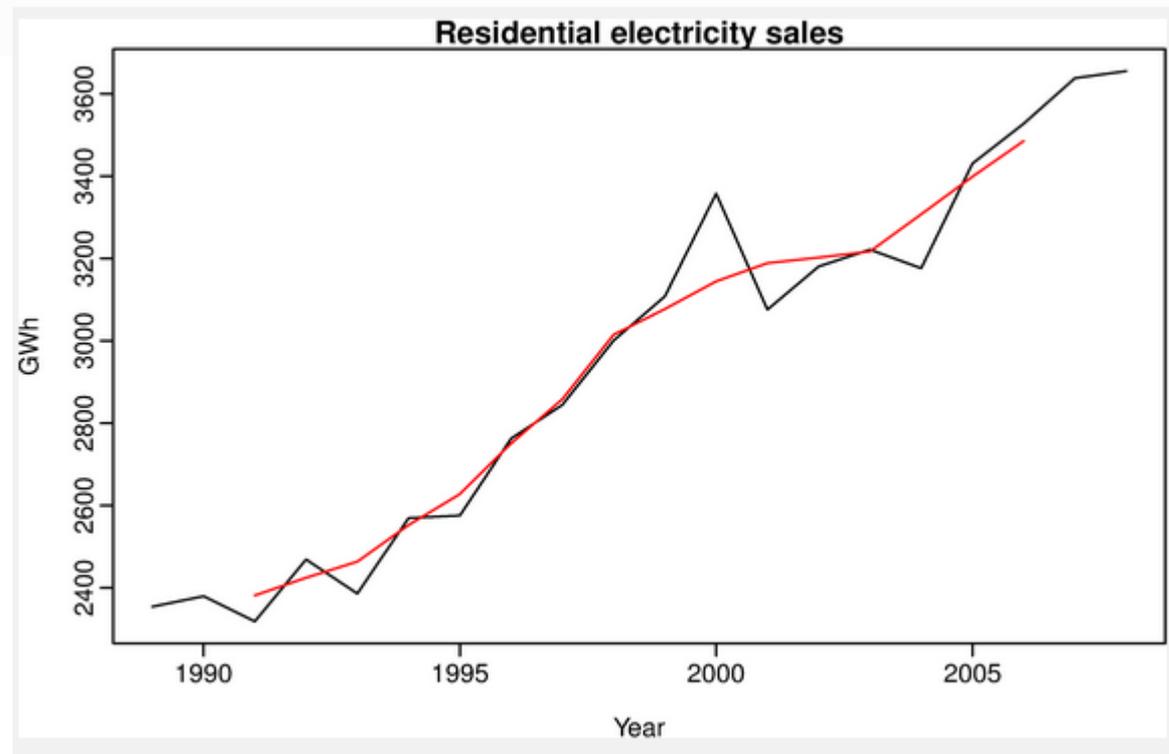
Adequate when the magnitude of the seasonal fluctuations or the variation around the trend-cycle does not vary with the level of the time series

Multiplicative decomposition

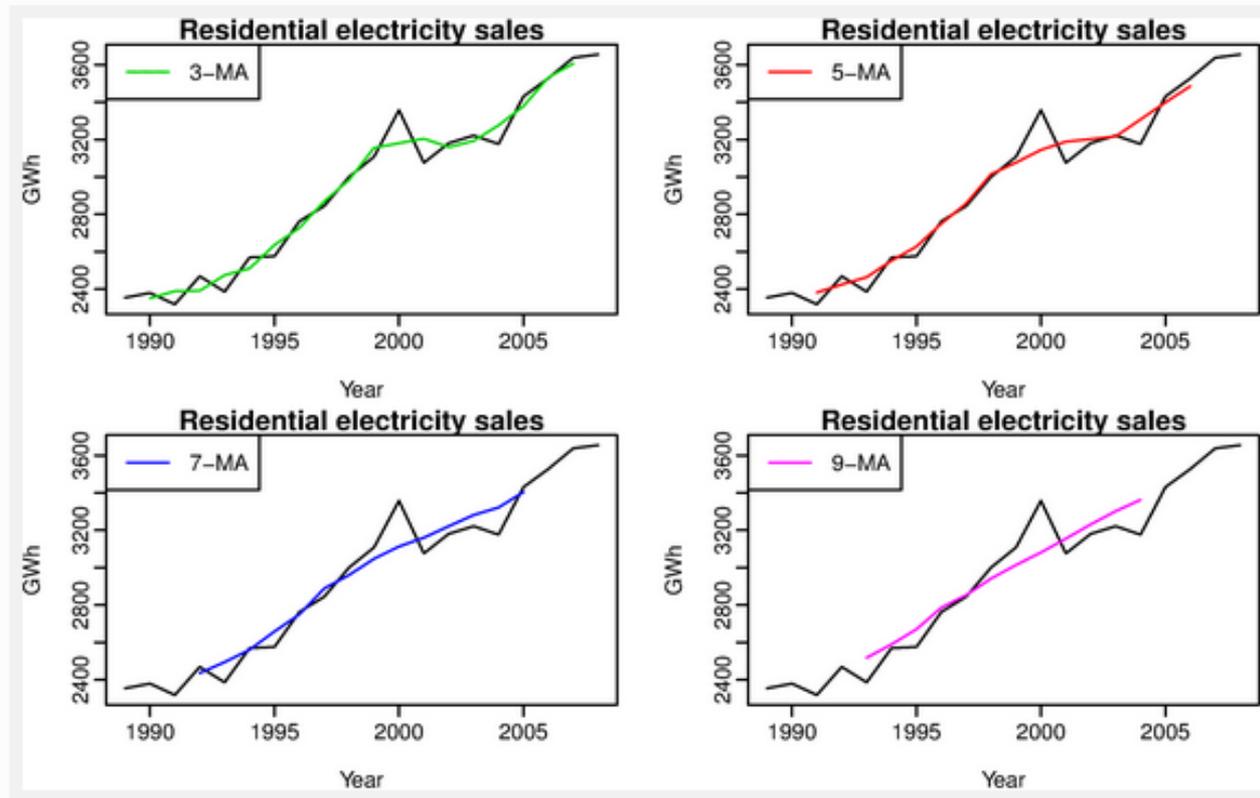
$$y_t = S_t \times T_t \times E_t$$

# Moving averages

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^k y_{t+j}$$



# Moving averages



Frequently used to estimate the trend-cycle from seasonal data

# STL decomposition

---

- STL is a robust and versatile decomposition method: Seasonal and Trend decomposition using Loess.
  - It can handle any type of seasonality
  - The seasonal component is allowed to change over time, within a range controllable by the user
  - The smoothness of the trend-cycle can also be controlled by the user
  - It is robust to outliers

# Forecasting with decomposition

---

- To forecast a decomposed time series, we forecast individual components, and then compute the predicted value

$$y_t = \hat{S}_t + \hat{A}_t$$

$$y_t = \hat{S}_t \hat{A}_t$$

# Contents

---

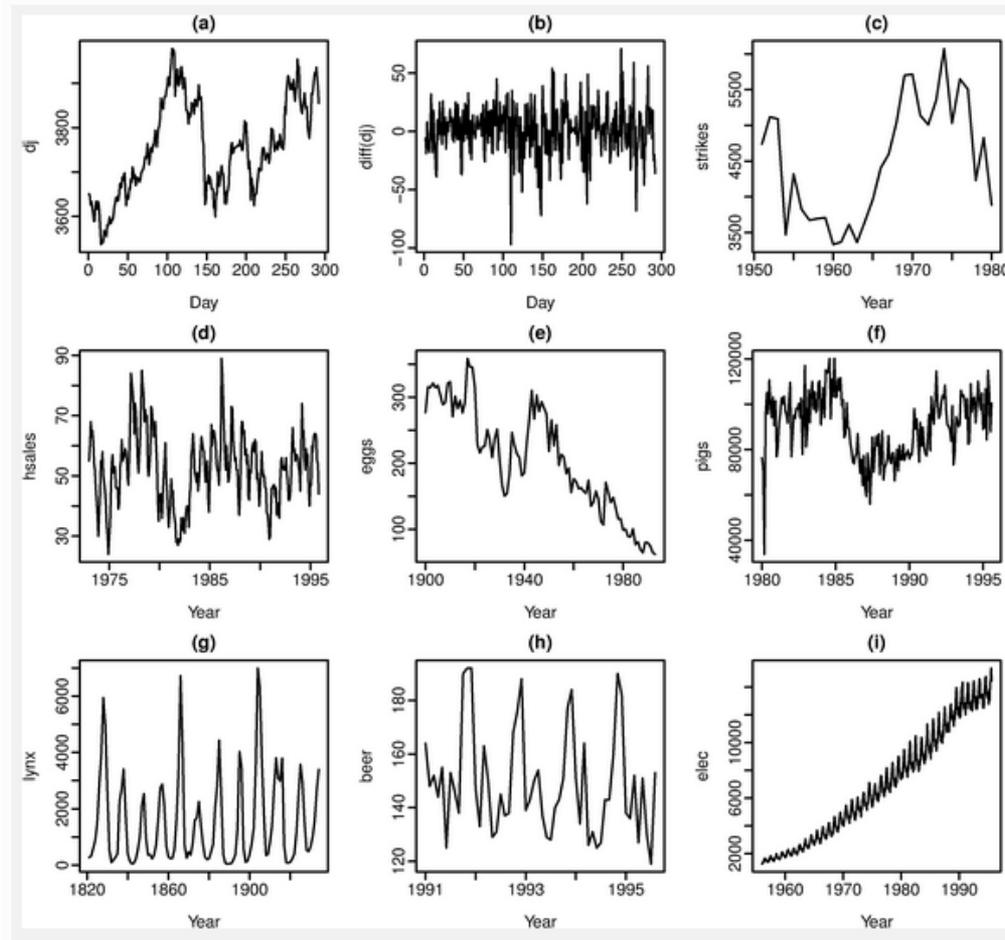
- Forecasting
- Forecaster's toolbox
- Simple regression
- Multivariate regression
- Time series decomposition

- **ARIMA models**

$$\hat{Y}_{t+1} = f(Y_t, Y_{t-1}, Y_{t-2}, \dots)$$

- Advanced forecasting models

# Example of time series



# Differencing

---

- Computing differences between successive observations
- Transformations such as logarithms can help to stabilize the variance of a time series. Differencing can help stabilize the mean of a time series by removing changes in the level of the time, and so eliminating trend and seasonality

# Random walk model

---

- A time series built by adding the error term to each new value:

$$y_t = y_{t-1} + e_t$$

- where the mean of  $e_t$  is zero and its sd is constant
- Random walks typically have:
  - long periods of apparent trends up or down
  - sudden and unpredictable changes in direction

# Unit root tests

---

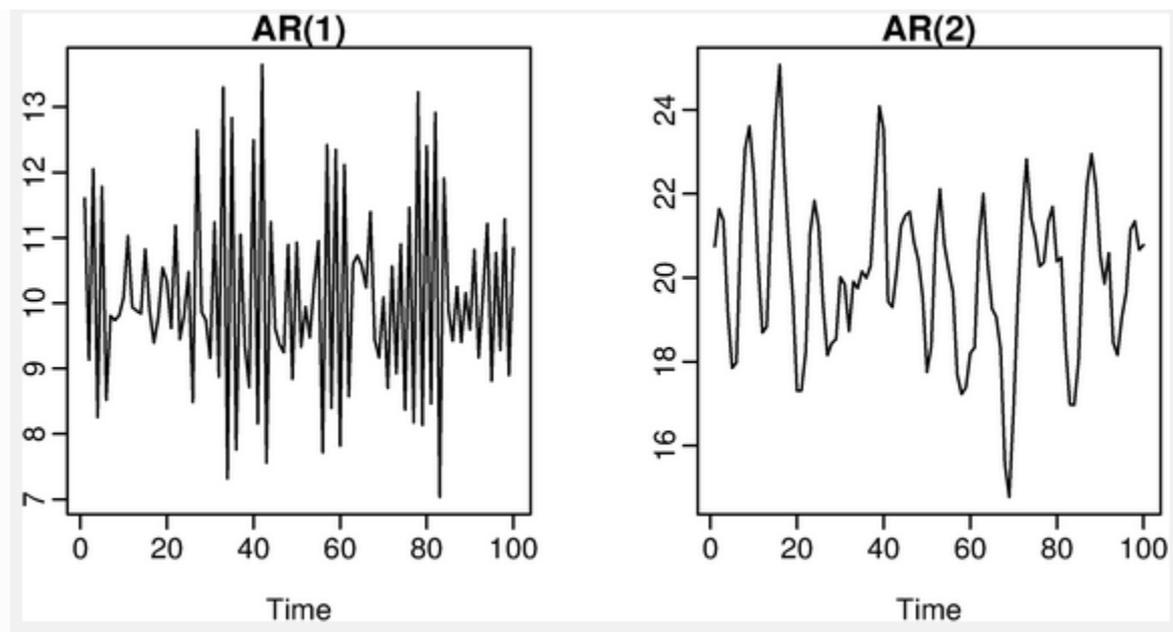
- Statistical hypothesis tests of stationarity designed for determining whether differencing is required
- Augmented Dickey-Fuller test

$$y'_t = \phi y_{t-1} + \beta_1 y'_{t-1} + \beta_2 y'_{t-2} + \cdots + \beta_k y'_{t-k}$$

# Autoregressive models

---

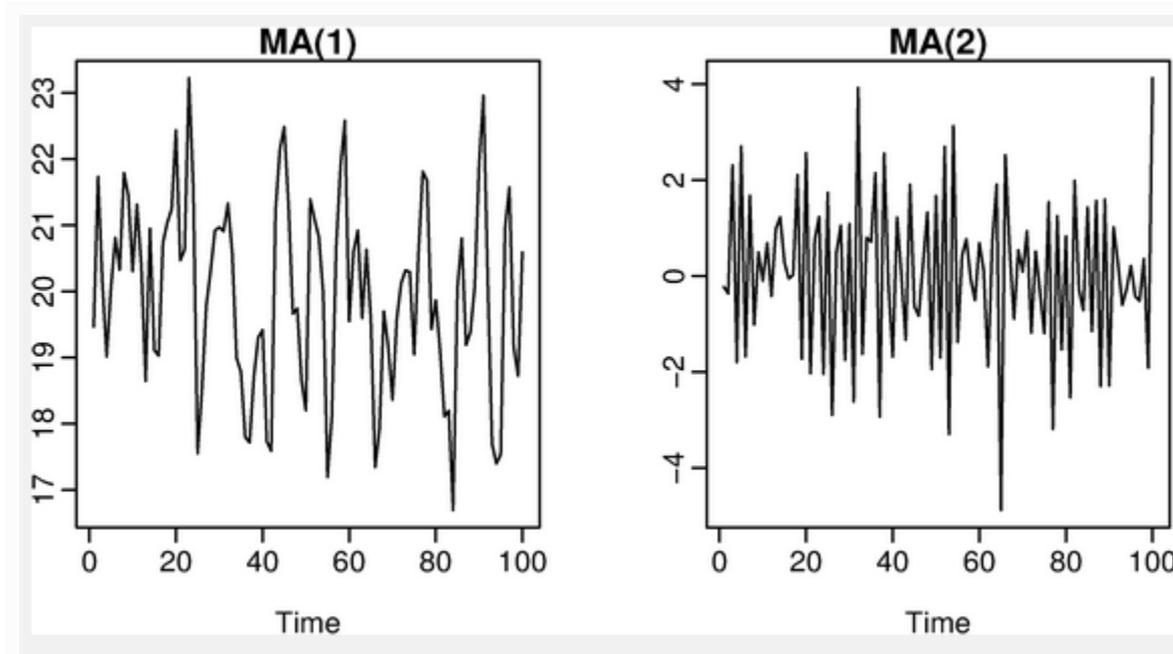
$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t$$



# Moving average models

---

$$y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$



# Non-seasonal ARIMA models

---

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t$$

- ARIMA(p,d,q)
  - p: order of the autoregressive part
  - d: degree of the first differencing part
  - q: order of the moving average part

White noise	ARIMA(0,0,0)
Random walk	ARIMA(0,1,0) with no constant
Random walk with drift	ARIMA(0,1,0) with a constant
Autoregression	ARIMA(p,0,0)
Moving average	ARIMA(0,0,q)

# Contents

---

- Forecasting
- Forecaster's toolbox
- Simple regression
- Multivariate regression
- Time series decomposition
- ARIMA models

$$\hat{Y}_{t+1} = f(Y_t, Y_{t-1}, Y_{t-2}, \dots)$$

- **Advanced forecasting models**

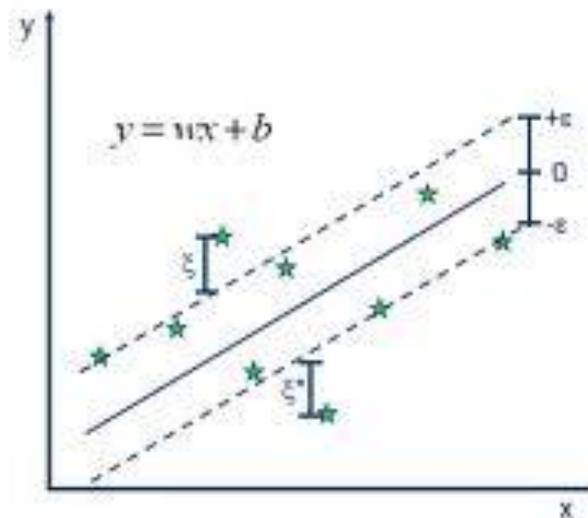
# Neural networks

---

- Multilayered perceptrons
- RBF
- Recurrent neural networks

# Support Vector Regression

- This the version of kernel machines for regression tasks



• Minimize:

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*)$$

• Constraints:

$$y_i - wx_i - b \leq \epsilon + \xi_i$$

$$wx_i + b - y_i \leq \epsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0$$

# References

---

- C. Chatfield, «The analysis of time series: An Introduction», Chapman & Hall/CRC, 2003
- J.D. Hamilton, «Time Series Analysis», Princeton University Press, 1994
- R. Hyndman, G. Athanasopoulos, «Forecasting and time series» 2013
- P.J. Brockwell, R.A. Davis, «Time Series: Theory and Methods», 2nd Ed., Springer, 1991
- J.S. Armstrong (ed), «Principles of Forecasting: A Handbook for Researchers and Practitioners», Springer, 2001

# References

---

- S.G. Makridakis, S.C. Wheelwright, R.J. Hyndman, «Forecasting», 3rd Ed., Wiley & Sons, 1998
- P.J. Brockwell, R.A. Davis, «Introduction to Time Series and Forecasting», 2nd ed., Springer, 2002
- A.K.Palit, D. Popovic, «Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications», Springer, 2005
- R.H. Shumway, D.S. Stoffer, «Time Series Analysis and Its Applications», Springer, 2nd Ed., 2006



# INTELIGENCIA DE NEGOCIO

2020 - 2021

- 
- Tema 1. Introducción a la Inteligencia de Negocio
  - Tema 2. Minería de Datos. Ciencia de Datos
  - Tema 3. Modelos de Predicción: Clasificación, regresión y series temporales
  - Tema 4. Preparación de Datos
  - Tema 5. Modelos de Agrupamiento o Segmentación
  - Tema 6. Modelos de Asociación
  - Tema 7. Modelos Avanzados de Minería de Datos.
  - Tema 8. Big Data